

П. Є. Пустовойтов

МАТЕМАТИЧНІ МОДЕЛІ МЕРЕЖ ЗВ'ЯЗКУ

Навчальний посібник

для студентів спеціальності
172 Телекомунікації та радіотехніка

РЕКОМЕНДОВАНО

Вченою радою університету
протокол № 4/10 від 29.03.2019 р.

Харків 2019

ББК
УДК

Упорядник: П.Є. Пустовойтов

Рецензенти:

Д.В. Агєєв, доктор техн. наук, ХНУРЕ

А.А. Серков, доктор техн. наук, НТУ «ХП»

Пустовойтов П.Є. Математичні моделі мереж зв'язку: Навчальний посібник / П.Є. Пустовойтов – Харків: ХНУРЕ, 2019. – 105 с.

Представлено математичні моделі вузлів мереж як системи масового обслуговування. Розглядаються моделі інтерфейсів вузлів мереж з однорідним і неоднорідним потоками, а також багатоканальних вузлів мереж з неоднорідним потоком без пріоритетів, з абсолютним і відносним пріоритетами

Призначено для студентів спеціальності 172 Телекомунікації та радіотехніка.

ISBN 978-966-659-271-5

© П. Є. Пустовойтов, 2019
©ХНУРЕ, 2019

Зміст

Вступ	5
1. Введення в системи масового обслуговування	7
1.1. Основні компоненти моделей масового обслуговування	7
1.2. Загальна модель системи масового обслуговування	8
1.3. Параметри систем масового обслуговування	13
1.4. Функціональні характеристики стаціонарних систем передачі пакетів.	15
Висновки.	20
Контрольні питання.	20
2. Моделі вузлів мережі з однорідним трафіком.	21
2.2 Розподіл часу очікування в моделі $(M / M / 1) : (FCFS / \infty / \infty)$.	25
2.3 Модель $(M / M / 1) : (GD / N / \infty)$.	27
2.5.3. Моделі з паралельними сервісами.	29
2.5.4. Модель $(M / M / R) : (GD / K / K)$ при $R < K$.	36
2.6. Модель $(M / G / 1) : (GD / \infty / \infty)$.	39
2.7. Моделі прийняття рішень в теорії масового обслуговування.	41
2.7.1. Модель з вартісними характеристиками	41
2.7.2. Модель кращого рівня обслуговування.	46
Висновки.	48
Контрольні питання.	48
3 Моделі вузлів мереж з неоднорідним трафіком з пріоритетами.	50
3.1. Інфокомунікаційні мережі, як системи масового обслуговування.	50
3.2. Модель інтерфейсу вузла мережі з неоднорідним вхідним потоком пакетів без переваг.	52
3.3. Модель інтерфейсу вузла мережі з неоднорідним вхідним потоком пакетів з абсолютним пріоритетом.	59
3.4. Модель багатоканального вузла мережі з неоднорідним вхідним потоком пакетів	66

<i>3.5. Модель вузла маршрутизації з неоднорідним вхідним потоком пакетів з відносним пріоритетом</i>	77
<i>3.6. Наближена технологія оцінки ефективності багатоканальної системи з відмовами і неоднорідним вхідним потоком з пріоритетами</i>	86
<i>Висновки.</i>	99
<i>Контрольні питання.</i>	99
ВИСНОВКИ	101
Список використаних джерел	103

Вступ

Традиційні технології аналізу, оцінки ефективності функціонування ТКС та вирішення задач управління їх ресурсами, засновані на використанні математичних моделей систем масового обслуговування (СМО) [1–3]. При цьому найбільш істотні результати теорії масового обслуговування (ТМО) пов'язані з використанням марківських моделей випадкового процесу [4–8], що описує вхідний потік, а також процеси обслуговування та очікування. Ці методи можуть бути успішно застосовані в припущенні про пуассонівський характер вхідного потоку.

Наявна подібність процесів функціонування мереж та систем масового обслуговування визначає можливість використання для аналізу мереж добре розвиненим математичним апаратом теорії масового обслуговування [1, 2]. Ця теорія спирається на потужний математичний апарат марківських випадкових процесів [4,6]. При цьому фактично, використовується допущення про те, що розподіл інтервалів між запитами та тривалістю їх обслуговування є експонентними. Відповідні математичні моделі обслуговування систем виявляються простими та елегантними. Крім того, використання припущення про посилення вхідного потоку запитів, а також потоку замовлених заяв задає нижні, песимістичні оцінки ефективності систем обслуговування (з усіх випадкових потоків із заданою інтенсивністю пуассонівський потік є самим випадковим). Ці обставини визначають і пояснюють широке застосування марківських моделей систем обслуговування. Використовуючи ці моделі, були отримані прості аналітичні співвідношення для розрахунку розподілів можливостей станів багатоканальних систем обслуговування з відхиленнями, з очікуванням при обмеженні по довжині черги та тривалості очікування в черзі, а також для систем без втрат.

Засобами стандартних марківських моделей ТМО вирішуються всі завдання аналізу одно– і багатоканальних систем, на вхід яких надходить однорідний потік запитів або суперпозиція потоків без пріоритетів [1]. Однак, у завданнях з пріоритетами повні результати відомі тільки в тому випадку, коли сумарний неоднорідний потік надходить в одноканальну систему. Складність вирішення цієї задачі для багатоканальної системи з неоднорідним потоком визначається швидким зростанням її розмірів. Так для звичайних ординарних потоків на вході та кількості каналів системи розмір завдання досягає десятків або сотень тисяч. Тому єдиний ефективний шлях вирішення таких завдань полягає у використанні методу фазового збільшення стану, суть якого полягає в декомпозиційному підході.

При оцінці продуктивності мережної інфраструктури, в першу чергу, слід розрізняти реальну завантаженість мережі, визначену за даними моніторингу трафіку, і пікову завантаженість. Під піковою завантаженістю слід розуміти завантаженість, що виникає в ситуації, коли всі робочі станції, використовуючи будь-які доступні види трафіку, обмінюються даними з усіма комп'ютерами та серверним устаткуванням. Якщо при розробці проекту модернізації мережі на основі даних, отриманих в результаті моделювання, орієнтуватися на пікові навантаження, то обсяг необхідних матеріальних засобів для модернізації буде надмірно великим. З упевненістю можна стверджувати, що така ситуація практично ніколи не з'явиться, але, за умови її моделювання, можливо виявити потенційно вузькі місця при розробці проекту мережної інфраструктури.

Найбільш природною є модель реальної завантаженості мережі. У цій моделі можна змінювати багато параметрів мережного обміну даними, що призводить до появи декількох варіантів вирішення поставленого завдання. Така модель дозволяє моделювати роботу мережі в динаміці, тобто оцінювати її ефективність залежно від поточного часу або доби. Це допоможе створити більш ефективні рішення щодо оптимізації архітектури мережі.

1. Введення в системи масового обслуговування

1.1. Основні компоненти моделей масового обслуговування

Основними елементами моделі масового обслуговування є клієнт (мережевий пакет, запит або вимога на обслуговування або просто об'єкт обслуговування) та сервіс (сервісний пристрій, мережний вузол, комутатор, засоби обслуговування тощо). Пакети подаються в систему обслуговування з деякого джерела. Надійшовши до мережі, вони можуть відразу потрапити на обслуговування або очікувати в черзі, якщо послуга зайнята. Після завершення процедури обслуговування сервіс автоматично вибирає з черги, якщо вона є, один з пакетів з тим, щоб розпочати його обслуговування. Якщо ж черга відсутня, то сервіс стає незайнятим до надходження нового пакета.

Надходження пакетів в систему обслуговування характеризується інтервалом між їх послідовними надходженнями, а обслуговування – часом обслуговування пакету. В загальному випадку ці параметри можуть бути і випадковими, і детермінованими, наприклад, як IP–трафік і АТМ – кадри.

В аналізі систем обслуговування визначну роль грає довжина черги, яка може бути кінцевою, як в буферній зоні між двома послідовними сервісними пристроями, та нескінченною, як у сервісного оператора поштових відділень.

Важливим фактором при аналізі систем обслуговування є дисципліна черги (або принцип побудови черги), що визначає порядок, відповідно до якого відбираються пакети з черги для обслуговування. Найбільш поширений принцип побудови черги, заснований на принципі «першим прийшов – першим буде обслуговуватися». Цей принцип найчастіше позначається аббревіатурою FIFO – від англійської First-In-First-Out, тобто «першим прийшов – першим вийшов». Друге правило, яке визначає принцип побудови черги – "останній прийшов – перший буде обслуговуватись", як правило, позначається як LIFO – від англ. Last-In-First-Out, тобто "останнім зайшов – першим вийшов". Потім потрібно вказати дисципліну черги, яка визначається випадковим правилом відбору клієнтів, іноді позначається як SIRO – від англійської Service-In-Random-Out, тобто "обслуговування по випадковому принципу". Крім того, пакети можуть відбиратися з черги відповідно до заданого пріоритету. Наприклад, при схемі обробки черг WFQ (Weighted Fair Queuing) та WRED (Weighted Random Early Detection).

При аналізі систем з чергами важливим фактором є також поведінка пакету, що потребує обслуговування. Він також може відмовитися від

очікування в черзі й покинути її, простояв в ньому деякий час, і за який дані в ньому застаріли.

Структура сервісної системи може включати в себе один сервіс або кілька таких засобів обслуговування, що працюють паралельно (багатоканальні та багатопроцесорні системи обслуговування). Крім того, сервіси можуть бути розташовані послідовно (наприклад, обслуговування представляє собою комплекс мережних вузлів, які пакет проходить послідовно для досягнення кінцевого вузла).

Джерело, яке генерує клієнтів, яких потрібно обслуговувати, може мати кінцеву або нескінченну потужність. Джерело кінцевої потужності обмежує кількість пакетів, що надходять на обслуговування (наприклад, передача єдиного відео-файлу). Навпаки, джерело нескінченної потужності завжди має вдосталь пакетів (наприклад, трансляція ТВ-програми).

Таким чином, можливо побудувати безліч моделей систем масового обслуговування, що моделюють роботу інформаційно-комунікаційних мереж, змінюючи перераховані вище операційні характеристики систем.

1.2. Загальна модель системи масового обслуговування

В розділі розглядаються загальні системи масового обслуговування, які мають і вхідний потік пакетів, і вихідний потік оброблених пакетів. Час між послідовними надходженнями пакетів та обслуговуванням часу є експоненційно розподіленими випадковими величинами. Ця модель служить основою при розгляді спеціалізованих моделей Пуассона [1].

При розгляді спільних систем масового обслуговування передбачається, що система функціонує протягом досить великого інтервалу часу, після закінчення якого в її роботі настає стаціонарний режим. Цей режим функціонування сервісної системи протиставляється перехідному (або невстановленому) режиму, який переважає в самий початковий період функціонування системи.

В розглянутій загальній моделі системи масового обслуговування передбачається, що інтенсивність надходження пакетів, а також інтенсивність вихідного потоку залежать від стану системи, що позначає їх залежність від кількості пакетів в системі обслуговування. Наприклад, тривала робота та перегрівання обладнання може знизити інтенсивність обслуговування, або, наприклад, довготривале виявлення пакета в черзі може привести до того, що джерело не досягне підтвердження доставки пакета і відправить його ще раз, тим самим збільшуючи інтенсивність вхідного трафіку.

Введемо наступні позначення.

n – кількість пакетів у системі обслуговування (в черзі та на обслуговуванні),

λ_n – інтенсивність надходження в систему пакетів за умови, що в системі вже є n пакетів,

μ_n – інтенсивність вихідного потоку оброблених пакетів за умови, що в системі знаходиться n пакетів,

p_n – ймовірність того, що в системі знаходиться n пакетів.

В загальній моделі системи масового обслуговування встановлюється функціональна залежність ймовірностей p_n від λ_n і μ_n . Ці ймовірності надалі використовуються при визначенні функціональних характеристик сервісної системи, таких як середня довжина черги, середня тривалість очікування пакетів обслуговування та середній коефіцієнт використання сервісів.

Ймовірності p_n визначаються з графу станів і переходів (діаграми інтенсивностей переходів), представленої на рис. 1.1. Сервісна система (вузол мережі) знаходиться в стані n , якщо в ньому є n пакетів. Ймовірність появи завбільшки одного нового пакета протягом досить малого проміжку часу h прагне до нуля при $h \rightarrow 0$. Це позначає, що при $n > 0$ стан n може бути змінено в двох можливих напрямках: у напрямку $n-1$, коли з інтенсивністю обслуговування оброблений пакет виходить з системи, і $(n+1)$ коли пакети надходять в систему з інтенсивністю λ_n . Стан 0 може змінитися тільки до стану 1, коли має місце надходження пакету з інтенсивністю λ_0 .

Відзначимо, що μ_0 не визначено, тому що пакети не можуть вийти з порожньої системи обслуговування.

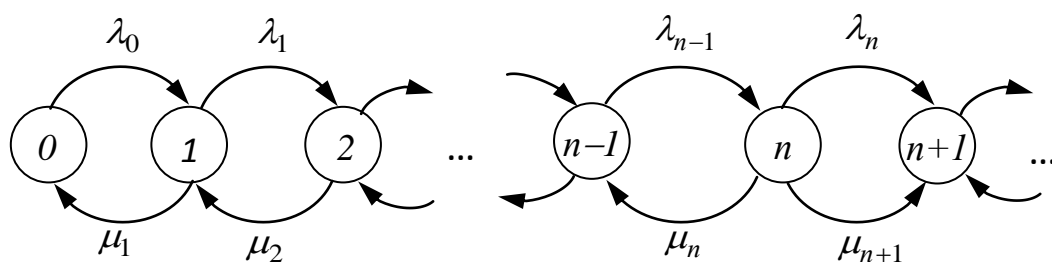


Рисунок 1.1 – Граф стану і переходів

При виконанні умов стаціонарності очікувані інтенсивності вхідного і вихідного потоків в стані $(n > 0)$ повинні бути рівними. Так як стан n може змінюватися тільки до стану $n-1$ і $n+1$, звідси випливає

$$\left(\begin{array}{c} \text{Очікувана інтенсивність} \\ \text{вхідного потоку в стані } n \end{array} \right) = \lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1}.$$

Аналогічно

$$\left(\begin{array}{c} \text{Очікувана інтенсивність} \\ \text{вихідного потоку в стані } n \end{array} \right) = (\lambda_n + \mu_n)P_n.$$

Дорівняв ці дві інтенсивності, отримуємо наступне рівняння балансу.

$$\lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} = (\lambda_n + \mu_n)P_n, n=1,2,\dots \quad (1.1)$$

Як видно з рис. 1.1, рівняння балансу (1.1), відповідне $n=0$, має вигляд

$$\lambda_0P_0 = \mu_1P_1. \quad (1.2)$$

Рівняння балансу вирішуються рекурентно. Вираз ймовірності p_i за допомогою p_0 послідовно змінюється наступним чином:

Для $n=0$ маємо

$$p_1 = \left(\frac{\lambda_0}{\mu_1} \right) p_0. \quad (1.3)$$

Для $n=1$ отримуємо

$$\lambda_0P_0 + \mu_2P_2 = (\lambda_1 + \mu_1)P_1. \quad (1.4)$$

Підставляючи сюди $p_1 = (\lambda_0/\mu_1)p_0$ та спрощуючи отриманий вираз, маємо

$$p_2 = \left(\frac{\lambda_1\lambda_0}{\mu_2\mu_1} \right) p_0. \quad (1.5)$$

Методом індукції можливо показати, що

$$p_n = \left(\frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_1} \right) p_0, n = 1, 2, \dots \quad (1.6)$$

Значення p_0 визначається з рівняння $\sum_{n=0}^{\infty} p_n = 1$.

Приклад 1.1

Система обслуговування довгих Data Mining запитів складається з трьох серверів. Принцип роботи системи обслуговування наступний: якщо кількість запитів менше чотирьох, то працюватиме лише один сервер. Якщо кількість запитів від чотирьох до шести, то будуть працювати два сервера. Якщо в наявності більше шести запитів, будуть працювати всі три сервера. Запит на обслуговування запитів надходять в систему до кас відповідно до розподілу Пуассона [1] з математичним очікуванням 10 осіб на годину.

Час виконання одного запиту сервера розподілено за експонентним законом із середнім значенням, що дорівнює 12 хвилинам. Визначимо у встановленому режимі ймовірність p_n , що n запитів стоять у черзі в систему.

З формулювання завдання маємо наступне

$\lambda_n = \lambda = 10$ запитів в годину, $n = 0, 1, \dots$,

$$\mu_n = \begin{cases} \frac{60}{12} = 5 \text{ запр. в год}, & n = 0, 1, 2, 3, \\ 2 \times 5 = 10 \text{ запр. в год}, & n = 4, 5, 6, \\ 3 \times 5 = 15 \text{ запр. в год}, & n = 7, 8, \dots \end{cases}$$

Відповідно,

$$p_1 = \left(\frac{10}{5} \right) p_0 = 2p_0,$$

$$p_4 = \left(\frac{10}{5} \right)^3 \left(\frac{10}{10} \right) p_0 = 8p_0,$$

$$p_2 = \left(\frac{10}{5} \right)^2 p_0 = 4p_0,$$

$$p_5 = \left(\frac{10}{5} \right)^3 \left(\frac{10}{10} \right)^2 p_0 = 8p_0,$$

$$p_3 = \left(\frac{10}{5}\right)^3 p_0 = 8p_0,$$

$$p_6 = \left(\frac{10}{5}\right)^3 \left(\frac{10}{10}\right)^3 p_0 = 8p_0,$$

$$p_n = \left(\frac{10}{5}\right)^3 \left(\frac{10}{10}\right)^3 \left(\frac{10}{15}\right)^{n-6} p_0 = 8\left(\frac{2}{3}\right)^{n-6} p_0, \quad n = 7, 8, \dots$$

Значення p_0 визначається з рівняння

$$p_0 + p_0 \left[2 + 4 + 8 + 8 + 8 + 8 + 8 + 8\left(\frac{2}{3}\right) + 8\left(\frac{2}{3}\right)^2 + 8\left(\frac{2}{3}\right)^3 + \dots \right] = 1$$

Або, що є рівноцінним,

$$p_0 \left\{ 31 + 8 \left[1 + \left(\frac{2}{3}\right) + \left(\frac{2}{3}\right)^2 + \dots \right] \right\} = 1.$$

Використовуючи формулу суми нескінченної геометричної прогресії

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}, \quad |x| < 1,$$

Отримуємо наступне

$$p_0 \left[31 + 8 \left(\frac{1}{1 - \frac{2}{3}} \right) \right] = 1.$$

Відповідно, $p_0 = 1/35$.

Якщо відома p_0 , можливо визначити будь-яку ймовірність, яка має відношення до завдання. Наприклад, ймовірність того, що буде працювати лише один сервер, обчислюється як ймовірність виявлення в системі не більше трьох запитів, тобто

$$p_0 + p_1 + p_2 + p_3 = (1 + 2 + 4 + 8)(1/35) \approx 0.273.$$

1.3. Параметри систем масового обслуговування

На рис. 1.2 схематично представлена спеціалізована система обслуговування пуассонівського типу, в якій паралельно функціонують ідентичні сервіси (засоби обслуговування). Вхідний пакет вибрано з черги для обслуговування на першому безкоштовному сервісі. Інтенсивність надходження пакетів в систему дорівнює λ пакетів в одиницю часу. Всі паралельні сервіси є ідентичними; це означає, що інтенсивність обслуговування кожного сервісу дорівнює μ пакетів в одиницю часу. Число пакетів, що знаходяться в системі обслуговування, включає пакети, які вже обслуговуються, і пакети, які знаходяться в черзі.

Позначення, що найбільш підходять для характеристик системи обслуговування (рис 1.2), мають наступну структуру:

$$(a/b/c):(d/e/f), \quad (1.7)$$

де

a – тип розподілу моментів часу надходження пакетів в систему,

b – тип розподілу часу між появою елементів вихідного потоку (часу обслуговування),

c – кількість сервісів, що працюють паралельно, $(=1, 2, \dots, \infty)$

d – дисципліна обслуговування черги,

e – максимальна ємність (кінцева або нескінченна) системи (кількість пакетів у черзі плюс кількість пакетів, прийнятих на обслуговування сервісів),

f – ємність (кінцева або нескінченна) джерела, що генерує пакети.



Рисунок 1.2 – Система обслуговування з декількома сервісами.

Стандартними позначеннями для типів розподілів вхідного та вихідного потоків (символи a і b) є наступні.

M – марківський (або пуассонівський) розподіл моментів надходження пакетів у систему або їх виходу з неї (або еквівалентний експонентний розподіл інтервалів часу між моментами послідовних подій або тривалості обслуговування пакетів),

D – детермінований (фіксований) інтервал часу між моментами послідовних надходжень в систему пакетів (або детермінована (фіксована) тривалість обслуговування пакетів),

E_k – розподіл Ерланга, або гамма-розподіл інтервалів часу (або, що те ж саме, розподіл суми незалежних випадкових величин, що мають експонентний розподіл),

GI – будь-який (загальний) тип розподілу моментів надходження пакетів на обслуговування,

GI – будь-який (загальний) тип розподілу тривалості обслуговування пакетів.

Для дисципліни обслуговування черги (символ d) використовуються наступні позначення:

$FCFS$ – першим прийшов – першого будуть обслуговувати

$LCFS$ – останній прийшов – першого будуть обслуговувати,

$SIRO$ – випадковий вибір пакетів з черги,

GD – будь-який (загальний) тип дисципліни обслуговування.

Для ілюстрації розглянемо структуру системи обслуговування, яка відповідає моделі $(M / D / 10) : (GD / N / \infty)$. Відповідно до прийнятих позначень, тут мова йде про систему (і, відповідно, моделі) масового обслуговування з пуассонівським вхідним потоком (або експонентним розподілом інтервалів часу між моментами послідовних надісланих пакетів), фіксованим часом обслуговування та десятьма сервісами, які функціонують паралельно. При цьому дисципліна черги не регламентована, а максимальна кількість допущених до системи пакетів дорівнює N . Насамкінець, джерело, яке генерує пакети, має необмежену ємність.

1.4. Функціональні характеристики стаціонарних систем передачі пакетів.

Детально розглядаючи систему обслуговування пуассонівського типу, покажемо що з допомогою отриманих в розділі 1.2 формул (1.3), (1.5) – (1.6) ймовірностей, які відповідають стаціонарному режиму, можливо отримати функціональні характеристики системи.

Основними функціональними характеристиками систем масового обслуговування є наступні [9 – 12]:

L_s – середнє число наявних в системі пакетів,

L_q – середнє число пакетів в черзі,

W_s – середня тривалість перебування пакета в системі,

W_q – середня тривалість перебування пакета в черзі,

\bar{c} – середня кількість зайнятих засобів обслуговування (сервісів).

Нагадаємо, що система включає як чергу, так і засоби обслуговування.

Покажемо, як перераховані функціональні характеристики отримано (прямо або побічно) з ймовірностей p_n – ймовірностей того, що в системі знаходиться n пакетів. Зокрема, маємо наступне.

$$L_s = \sum_{n=1}^{\infty} n p_n, \quad (1.8)$$

$$L_q = \sum_{n=c+1}^{\infty} (n - c) p_n. \quad (1.9)$$

Залежність між L_s і W_q (а також між L_q і W_q), що відома в літературі по теорії масового обслуговування як формула Літтла, має вигляд

$$L_s = \lambda_{ef} W_s, \quad (1.10)$$

$$L_q = \lambda_{ef} W_q. \quad (1.11)$$

Ці співвідношення справедливі при досить загальних умовах. Параметр λ_{ef} представляє собою ефективну інтенсивність надходження пакетів у систему обслуговування. Він дорівнює (вихідній) інтенсивності надходження клієнтів λ , коли всі надіслані пакети мають можливість потрапити до сервісної системи. Якщо деякі пакети не мають такої можливості з тієї причини, що система заповнена, тоді $\lambda_{ef} < \lambda$. Пізніше буде показано, як обчислюється λ_{ef} .

Існує також пряма залежність між значеннями W_s і W_q . За визначенням

$$\left(\begin{array}{c} \text{Середня тривалість} \\ \text{знаходження в системі} \end{array} \right) = \left(\begin{array}{c} \text{Середній час} \\ \text{перебування в черзі} \end{array} \right) + \left(\begin{array}{c} \text{Середній час} \\ \text{обслуговування} \end{array} \right)$$

Математично це записується в наступному вигляді

$$W_s = W_q + \frac{1}{\mu}. \quad (1.12)$$

Відтак можливо отримати формулу, яка пов'язує L_s та L_q , помножуючи обидві частини останнього співвідношення на λ_{ef} і використовуючи формулу Літтла. В результаті отримуємо

$$L_s = L_q + \frac{\lambda_{ef}}{\mu}. \quad (1.13)$$

За визначенням різниця між середнім числом клієнтів L_s , що знаходяться в системі, і середнім числом клієнтів у черзі L_q дорівнює середній кількості зайнятих вузлів обслуговування \bar{c} . Відповідно маємо

$$\bar{c} = L_s - L_q = \frac{\lambda_{ef}}{\mu}. \quad (1.14)$$

Тому коефіцієнт використання вузлів обслуговування вираховується як відношення

$$k_{вик} = c / \bar{c}. \quad (1.15)$$

Приклад 1.2

У базах даних існує п'ять диспетчерів, що обслуговують сесії з клієнтами. Запити до базу даних надходять відповідно до розподілу Пуассона з інтенсивністю в шість запитів у хвилину. Час підпорядкування сесії з клієнтом є експоненційно розподіленою випадковою величиною з середнім значенням, що дорівнює 30 сек. Клієнти бази даних, які не можуть знайти вільного диспетчера обслуговування, можуть тимчасово очікувати звільнення диспетчера в черзі.

В черзі може знаходитися не більше трьох клієнтів. Якщо черга зайнята, то новим клієнтам буде відмовлено в доступі. Потрібно визначити наступне:

- а) ймовірність p_n того, що в системі знаходиться n клієнтів,
- б) ефективну інтенсивність надходження клієнтів в систему,
- в) середню кількість клієнтів в черзі,
- г) середню тривалість очікування клієнта в черзі,
- д) середню кількість зайнятих диспетчерів бази даних.

Перш за все, відзначимо, що диспетчер баз даних в розглянутій ситуації виступає в ролі сервісу, таким чином система має лише $c=5$ засобів обслуговування. Максимальна місткість системи дорівнює $5+3=8$ клієнтських запитів.

Ймовірність може бути визначена як окремий випадок з загальної моделі, яка була розглянута у розділі 1.2, формула (1.6). Наприклад, маємо

$$\lambda_n = 6 \text{ запитів у хвилину, } n = 0, 1, 2, \dots, 8,$$

$$\mu_n = \begin{cases} n \left(\frac{60}{30} \right) = 2n \text{ запитів у хвилину,} & n = 1, 2, \dots, 5, \\ 5 \left(\frac{60}{30} \right) = 10 \text{ запитів у хвилину,} & n = 6, 7, 8. \end{cases}$$

Відповідно, використовуючи співвідношення (1.3), (1.5) – (1.6), обчислюємо

$$p_n = \begin{cases} \frac{\left(\frac{6}{2}\right)^n}{n!} p_0, & n = 1, 2, \dots, 5, \\ \frac{\left(\frac{60}{30}\right)}{5!5^{n-5}} p_0, & n = 6, 7, 8. \end{cases}$$

Значення p_0 обчислюємо шляхом підстановки значень для p_n , $n = 1, 2, \dots, 8$ в рівняння $p_0 + p_1 + \dots + p_8 = 1$. В результаті отримуємо

$$p_0 + p_0 \left(\frac{3}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!} + \frac{3^6}{5!5^1} + \frac{3^7}{5!5^2} + \frac{3^8}{5!5^3} \right) = 1.$$

Рішенням цього рівняння є $p_0 = 0,04812$. Знайдене значення p_0 дозволяє обчислити всі ймовірності від p_1 до p_8 .

Таблиця 1.1 – Статистичний ряд для прикладу 1.2

n	1	2	3	4	5	6	7	8
p_n	0,4436	0,21654	0,21654	0,16240	0,09744	0,05847	0,03508	0,02105

Ефективну інтенсивність надходження запитів λ_{ef} до системи можливо вираховувати з використанням принципової схеми (рис 1.3), відповідно до якої клієнтські запити з джерела надходять з інтенсивністю λ . Вхідний запит може потрапити до системи з інтенсивністю λ_{ef} або бути відфільтрованим з інтенсивністю $\lambda_{втррати}$, тобто $\lambda = \lambda_{ef} + \lambda_{втррати}$. Система не зможе обслуговувати запит клієнта, якщо там вже є 8 запитів. Це позначає, що частина клієнтів, які не зможуть потрапити в систему, є пропорційною p_8 . Отже,

$$\lambda_{втррати} = \lambda p_8 = 6 \times 0,02105 = 0,1263 \text{ запитів в хвилину,}$$

$$\lambda_{ef} = \lambda - \lambda_{втрати} = 6 - 0,1263 = 5,8737 \text{ запитів в хвилину.}$$

Середня кількість клієнтів в системі (клієнти, яких займають диспетчери, та клієнти, які очікують обслуговування в черзі) визначається значенням L_s – середнім числом клієнтів в системі. Значення L_s визначається за допомогою p_n наступний чином.

$$L_s = 0p_0 + 1p_1 + \dots + 8p_8 = 3,1286 \text{ клієнтських запитів.}$$

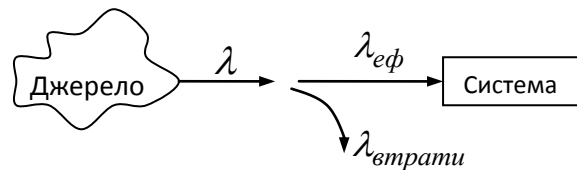


Рис. 1.3. Співвідношення між різними показниками інтенсивності.

Середня тривалість очікування клієнтського запиту в черзі дорівнює величині W_q . Для розрахунку W_q використовуємо співвідношення (1.12)

Оскільки

$$W_s = \frac{L_s}{\lambda_{ef}} = \frac{3,1286}{5,8737} = 0,53265 \text{ хвилини,}$$

тоді

$$W_q = 0,53265 - \frac{1}{2} = 0,03265 \text{ хвилини.}$$

Середня кількість зайнятих диспетчерів баз даних дорівнює середньому значенню зайнятих сервісів і тому обчислюється таким чином.

$$\bar{c} = L_s - L_q = \frac{\lambda_{ef}}{\mu} = \frac{5,8737}{2} = 2,9368 \text{ диспетчерів.}$$

Звідси отримуємо, що коефіцієнт використання диспетчерів дорівнює

$$k_{вик} = \frac{\bar{c}}{c} = \frac{2,9368}{5} = 0,58736.$$

Висновки.

Комп'ютерна мережа вирішує сукупність задач з обслуговування потоків пакетів, яким у теорії масового обслуговування відповідають заявки. Заявки генеруються різними джерелами випадковим чином або в випадкові моменти часу і надходять на вхід вузла–приймача комп'ютерної мережі. Потім вузлом–приймачем вони передаються вільному інтелектуальному вузлу для ініціалізації процесу обслуговування отриманої заявки. Якщо все інтелектуальні вузли в даний момент зайняті, то знову прибула заявка з вузла приймача надходить в чергу для очікування обслуговування. Для опису та вирішення таких задач успішно використовується математичний апарат систем масового обслуговування.

У розділі було розглянуто основи систем масового обслуговування, надано схему, яка описує процеси системи. Вирішено задачу розрахунку ймовірностей станів системи для випадку дискретного випадкового процесу із дискретною множиною станів у стаціонарному режимі. Надано формули для розрахунку загальних показників системи масового обслуговування та обґрунтована їх відповідність до показників якості обслуговування у мережах (QoS).

Контрольні питання.

1. Нарисуйте схему функціонування системи масового обслуговування, та вкажіть загальні вхідні та вихідні дані.
2. Поясніть параметри маркування систем масового обслуговування.
3. Перелічіть загальні характеристики стаціонарних систем та вкажіть формули.
4. Поясніть формулу Літтла.
5. В чому полягає мета вирішення системи масового обслуговування?
6. Яким показникам якості обслуговування у мережах (QoS), відповідають певні характеристики систем масового обслуговування?
7. Із якою метою використовується показник ефективного навантаження? Чому він дорівнює у системах із безкінечною чергою.
8. Що таке приведена інтенсивність вхідного потоку?
9. Що таке стаціонарний режим функціонування системи?
10. Чим стаціонарна система відрізняється від нестаціонарної? Як ця властивість відображена у формулах розрахунку ймовірностей станів?

2. Моделі вузлів мережі з однорідним трафіком.

В цьому розділі представлені дві моделі сервісної системи з одним засобом обслуговування (тобто $c=1$). Передбачається, що пакети поступають до вузла мережі з постійною інтенсивністю λ . Інтенсивність обслуговування також постійна і дорівнює μ пакетам в одиницю часу. Перша модель не встановлює обмежень на місткість системи, у другій моделі передбачається, що місткість системи обмежена. У цих двох моделях джерело клієнтів має необмежену ємність.

Будемо використовувати позначення Кендалла для опису параметрів системи обслуговування в кожному випадку. Завдяки тому, що виведення співвідношення (1.6) для розділу 1.2 і всіх функціональних характеристик сервісної системи (1.8) — (1.15) в розділі 1.4 виконано незалежно від конкретної дисципліни черги, в позначеннях ми будемо використовувати символ GD (дисципліна черги не регламентована).

2.1 Модель $(M / M / 1) : (GD / \infty / \infty)$.

Використовуючи позначення загальної моделі, маємо $\lambda_n = \lambda$ і $\mu_n = \mu$ для всіх $n = 0, 1, 2, \dots$

Оскільки відсутні обмеження на ємність черги вузла мережі, й, отже, всі наступні пакети можуть потрапити в систему обслуговування, то $\lambda_{ef} = \lambda$ і $\lambda_{втрати} = 0$.

Позначимо наведену інтенсивність вхідного потоку $\rho = \lambda / \mu$. Тоді вираз для ймовірності стану системи p_n в загальній моделі приймає наступний вигляд:

$$p_n = \rho^n p_0, \quad n = 0, 1, 2, \dots \quad (2.1)$$

Для визначення величини p_0 використовується твердження, що сума всіх ймовірностей дорівнює одиниці.

$$p_0(1 + \rho + \rho^2 + \dots) = 1. \quad (2.2)$$

Передбачаємо, що $\rho < 1$, тоді геометричний ряд має кінцеву суму $1/(1-\rho)$, тому

$$p_0 = 1 - \rho. \quad (2.3)$$

Відповідно, загальна формула для p_n має вигляд

$$p_n = (1 - \rho)\rho^n, \quad n = 1, 2, \dots, (\rho < 1). \quad (2.4)$$

Ці значення ймовірностей p_n (враховуючи і ймовірність p_0) відповідають ряду, який задається геометричною прогресією.

При виведенні формули (2.4) з самого початку передбачалося, що $\rho < 1$. Це означає, що для досягнення системою стаціонарного режиму функціонування необхідно, щоб інтенсивність надходження пакетів λ була строго менше інтенсивності обслуговування μ . В разі виконання умови, що $\lambda \geq \mu$, ряд геометричної прогресії має бути розбіжним, і, отже, ймовірності p_n стаціонарного стану не існували. В такому випадку система обслуговування буде функціонувати в нестаціонарному режимі, коли довжина черги з часом необмежено зростає.

Середнє число наявних в системі пакетів L_s як функціональна характеристика сервісної системи вузла мережі визначається за наступною формулою:

$$L_s = \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = (1-\rho)\rho \frac{d}{d\rho} \sum_{n=0}^{\infty} \rho^n = (1-\rho)\rho \frac{d}{d\rho} \left(\frac{1}{1-\rho} \right) = \frac{\rho}{1-\rho}. \quad (2.5)$$

Якщо в розглянутій моделі $\lambda_{ef} = \lambda$, тоді інші функціональні характеристики сервісної системи розраховуються з використанням співвідношень (1.4) – (1.11), що приводить до наступних результатів

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}, \quad (2.6)$$

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}, \quad (2.7)$$

$$L_q = \lambda W_q = \frac{\rho^2}{1 - \rho}, \quad (2.8)$$

$$\bar{c} = L_s - L_q = \rho. \quad (2.9)$$

Приклад 2.1

Маршрутизатор має тільки один сервісний процесор. Пакети надходять відповідно до розподілу Пуассона з середнім значенням 4 пакета на хвилину і можуть очікувати обслуговування в черзі. Час обслуговування пакета є експоненційно розподіленою випадковою величиною з математичним очікуванням 10 сек. Довжина черги в цьому випадку не обмежена. Необхідно визначити розмір буфера пам'яті для черги.

Для розглянутого прикладу ми маємо $\lambda = 4$ пакета в хвилину і $\mu = 60/10 = 6$ пакетів у хвилину. Так як $\rho = \lambda / \mu < 1$, то система може функціонувати в стаціонарному режимі. Ємність системи дорівнює ∞ і ємність джерела також дорівнює ∞ .

Основні параметри системи розраховуються за формулою (2.5) – (2.9).

$$\lambda_{ef} = 4, L_s = 2, W_s = 0.5, \frac{\rho}{c} = \rho = 0,66667, L_q = 1,3333, W_q = 0,3333.$$

Таким чином, отримуємо $L_q = 1,33$ пакета. Некоректно розглядати L_q як єдиний аргумент при визначенні шуканого місця в буфері, так як при розрахунку повинна враховуватися максимально можлива довжина черги. Наприклад, можна розрахувати максимально можливу кількість пакетів у пам'яті, при якій, якнайменше, 90% вхідних пакетів попадають в систему.

Таблиця 2.1 – Результати розрахунків для прикладу 2.1

n	p_n	P_n	n	p_n	P_n
0	0,33333	0,33333	13	0,00171	0,99657
1	0,22222	0,55556	14	0,00114	0,99772
2	0,14815	0,70370	15	0,00076	0,99848
3	0,09877	0,80247	16	0,00051	0,99899
4	0,06584	0,86831	17	0,00034	0,99932
5	0,04390	0,91221	18	0,00023	0,99955

6	0,02926	0,94147	19	0,00015	0,99970
7	0,01951	0,96098	20	0,00010	0,99980
8	0,01301	0,97399	21	0,00007	0,99987
9	0,00867	0,98266	22	0,00004	0,99991
10	0,00578	0,98844	23	0,00003	0,99994
11	0,00385	0,99229	24	0,00002	0,99996
12	0,00257	0,99486	25	0,00001	0,99997

Нехай невідома змінна s представляє шукану кількість пакетів в пам'яті. Тоді система має ємність $s + 1$.

Надісланий пакет в 90% випадків отримає місце в черзі, якщо в системі знаходиться максимум s пакетів. Ця умова еквівалентна наступному ймовірнісному твердженню:

$$p_0 + p_1 + \dots + p_s \geq 0,9.$$

З попередньої таблиці випливає, що суми ймовірностей p_n дорівнюють 0,86831 і 0,91221 при $n=4$ і $n=5$ відповідно. Це позначає, що умова виконується при $s > 5$ місць в черзі.

Кількість можливих пакетів в черзі s може бути також визначена з використанням формули, яка визначає p_n . Отримуємо

$$(1 - \rho)(1 + \rho + \rho^2 + \dots + \rho^s) \geq 0,9.$$

Сума усіченого геометричного ряду дорівнює $(1 - \rho^{s+1}) / (1 - \rho)$. Відповідно, останній вираз приводиться до вигляду

$$(1 - \rho^{s+1}) \geq 0,9.$$

Спрощуючи цю нерівність, отримуємо

$$\rho^{s+1} \leq 0,1.$$

Логарифмуючи обидві частини останньої нерівності, отримуємо наступне

$$s \geq \frac{\ln(0,1)}{\ln\left(\frac{4}{6}\right)} - 1 = 4,679 \approx 5.$$

Таким чином, розмір черги повинен бути таким, щоб вміщувалось $s \geq 5$ пакетів.

2.2 Розподіл часу очікування в моделі $(M / M / 1) : (FCFS / \infty / \infty)$.

Висновок формули (2.1) – (2.3) для ймовірностей p_n в загальній моделі системи обслуговування абсолютно не залежить від дисципліни черги. Це позначає, що математичні очікування всіх функціональних параметрів W_s , W_q , L_s и L_q справедливі для системи з будь-якої дисципліною черги.

На відміну від середнього часу очікування в системі обслуговування, щільність ймовірності його розподілу залежить від дисципліни черги. Проілюструємо це твердження шляхом побудови щільності ймовірності часу очікування для моделі $(M / M / 1)$ з дисципліною черги $FCFS$ ("першим прийшов – першим буде обслуговуватися").

Позначимо через τ кількість часу, яку щойно надісланий пакет проведе в системі від моменту прибуття до завершення обслуговування. Виходячи з дисципліни черги $FCFS$, якщо в системі вже перебуває n пакетів, які надійшли в систему перед щойно надісланим, тоді

$$\tau = t'_1 + t_2 + \dots + t_{n+1},$$

де t'_1 – час, необхідний для завершення обслуговування пакета, який вже знаходиться в засобі обслуговування системи, а t_2, t_3, \dots, t_n – інтервали часу, які будуть потрібні для обслуговування $n-1$ пакетів, які знаходяться в черзі. Величина t_{n+1} представляє собою час обслуговування щойно надісланого пакета.

Позначимо через $w(\tau | n+1)$ умовну щільність ймовірності τ , де умовою служить наявність в сервісній системі n пакетів на момент прибуття нового. Оскільки час обслуговування в системі розподілено по експонентному закону, який має властивість відсутності післядії, величина t'_1 також розподілена за експонентним законом. Отже, τ є сумою $n+1$ незалежних випадкових величин, кожна з яких підкоряється одному і тому експонентному розподілу. Як відомо з теорії ймовірностей, в цьому випадку функція $w(\tau | n+1)$ буде

щільністю ймовірності гамма-розподілу з параметрами μ і $n+1$. Звідси випливає, що

$$\begin{aligned} w(\tau) &= \sum_{n=0}^{\infty} w(\tau | n+1) p_n = \sum_{n=0}^{\infty} \frac{\mu(\mu\tau)^n e^{-\mu\tau}}{n!} (1-\rho) \rho^n = \\ &= (1-\rho) \mu e^{-\mu\tau} \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n}{n!} = \mu(1-\rho) e^{-\mu(1-\rho)\tau}, \quad \tau > 0. \end{aligned}$$

Таким чином, показано, що випадкова величина τ має експонентний розподіл з математичним очікуванням

$$W_s = 1 / \mu(1-\rho). \quad (2.10)$$

Приклад 2.2

В моделі роботи маршрутизатора з прикладу 2.1 цілком обґрунтованим є припущення про те, що обслуговування виконується відповідно до дисципліни черги *FCFS*. Визначимо надійність використання величини W_s в якості оцінки часу очікування в системі. Для цього оцінимо частину пакетів, час очікування яких перевищує W_s . Використовуючи формулу (2.10), отримуємо

$$P\{\tau > W_s\} = 1 - \int_0^{W_s} w(\tau) d\tau = e^{-\mu(1-\rho)W_s} = e^{-1} \approx 0,368.$$

Отже, при дисципліні черги *FCFS* для 37% пакетів час очікування перевищить значення W_s . Цей відсоток здається надмірно великим, особливо якщо врахувати, що поточне значення W_s для розглянутої системи обслуговування також є великим (0,5 хв). Зауважимо, що знайдена ймовірність ($e^{-1} \approx 0,368$) не залежить від інтенсивностей λ і μ моделі $(M / M / 1) : (FCFS / \infty / \infty)$, це в свою чергу позначає, що її величину неможливо зменшити. Отже, якщо проектується система на підставі середньої величини W_s , то слід очікувати, що для 36,8% пакетів час очікування перевищить середній час очікування в системі.

Існує дві можливості поліпшити ситуацію: 1) збільшити інтенсивність обслуговування пакетів μ для зменшення значення W_s до прийнятного рівня

або 2) вибрати інтенсивність обслуговування клієнтів таким чином, щоб ймовірність того, що час очікування перевищить заздалегідь визначену величину (скажімо, 10 сек), була менше прийнятно малої величини (наприклад, 10%). Перша можливість еквівалентна пошуку такого значення μ , що $W_s < \bar{T}$, а друга – обчисленню такого значення μ , що $P\{\tau > \bar{T}\} < \alpha$, де \bar{T} і α повинні визначатися користувачем.

2.3 Модель $(M / M / 1) : (GD / N / \infty)$.

Ця модель відрізняється від розглянутої вище тільки тим, що система вміщує не більше N пакетів (максимальна довжина черги дорівнює $(N - 1)$).

Ситуація в цій моделі така що, як тільки число пакетів в системі досягає N , жоден з додаткових пакетів на обслуговування не приймається. З цієї умови випливає, що

$$\lambda_n = \begin{cases} \lambda, & n = 0, 1, \dots, N-1, \\ 0, & n = N, N+1, \dots, \end{cases} \quad \mu_n = \mu, \quad n = 0, 1, \dots$$

Використовуючи позначення $\rho = \lambda / \mu$, маємо

$$p_n = \begin{cases} \rho^n p_0, & n \leq N, \\ 0, & n > N. \end{cases} \quad (2.11)$$

Значення ймовірності p_0 визначається з рівняння $\sum_{n=0}^x p_n = 1$, яке приймає наступний вигляд:

$$p_0(1 + \rho + \rho^2 + \dots + \rho^N) = 1.$$

Звідси отримуємо

$$p_0 = \begin{cases} \frac{1 - \rho}{1 - \rho^{N+1}}, & \rho \neq 1, \\ \frac{1}{N+1}, & \rho = 1. \end{cases} \quad (2.12)$$

Відповідно,

$$p_n = \begin{cases} \frac{(1-\rho)\rho^n}{1-\rho^{N+1}}, & \rho \neq 1, \\ \frac{1}{N+1}, & \rho = 1, \end{cases} \quad n = 0, 1, \dots, N. \quad (2.13)$$

Зауважимо, що в цій моделі значення параметру $\rho = \lambda / \mu$ не обов'язково має бути меншим одиниці тому, що надходження пакетів в систему контролюється максимальним обсягом системи N . Це означає, що в даному випадку в якості інтенсивності надходження пакетів швидше виступає λ_{ef} , ніж λ . Пакети можуть бути втраченими в тому випадку, якщо в системі знаходиться N пакетів, саме, як показано на рис. 1.3,

$$\lambda_{втрати} = \lambda p_N, \quad (2.14)$$

$$\lambda_{ef} = \lambda - \lambda_{втрати} = \lambda(1 - p_N). \quad (2.15)$$

Тоді слід очікувати, що λ_{ef} буде меншим за μ .

Середня кількість клієнтів в системі обчислюється за формулою

$$\begin{aligned} L_s &= \sum_{n=0}^N n p_n = \frac{1-\rho}{1-\rho^{N+1}} \sum_{n=0}^N n \rho^n = \frac{(1-\rho)\rho}{1-\rho^{N+1}} \frac{d}{d\rho} \left(\frac{1-\rho^{N+1}}{1-\rho} \right) = \\ &= \frac{\rho \{1 - (N+1)\rho^N + N\rho^{N+1}\}}{(1-\rho)(1-\rho^{N+1})}, \quad \rho \neq 1. \end{aligned} \quad (2.16)$$

При $\rho = 1$ виходить $L_s = 1$. Використовуючи значення L_s та λ_{ef} , можливо також отримати вирази для W_s , W_q та L_q .

Приклад 2.3

Розглянемо ситуацію з маршрутизатором з прикладу 2.1. Нехай вузол мережі має буфер пам'яті на чотири пакети. Якщо в пам'яті немає вільного місця, знов надіслані пакети будуть відфільтровані. Адміністратор мережі хоче визначити вплив обмеженого обсягу черги на втрати пакетів.

В позначеннях, прийнятих в моделі, максимальна місткість системи дорівнює $N = 4 + 1 = 5$.

Оскільки ємність системи дорівнює $N = 5$, частка втрачених пакетів становить $p_5 = 0,04812$. Рішення щодо збільшення обсягу буфера вузла мережі повинно ґрунтуватися на вимогах до мережі.

Аналізуючи ситуацію з іншого боку, помічаємо, що середній час перебування пакета в сервісній системі $W_s = 0,3736$ (приблизно 22 сек), тобто менше 30 сек, як це було в прикладі 2.2, коли всім вхідним пакетам дозволялося встати в чергу. Зменшення цього показника сервісної системи приблизно на 25% забезпечено за рахунок втрати близько 4.8% пакетів через обмежений обсяг пам'яті в вузлі мережі.

$$\begin{aligned}\lambda &= 4, \lambda_{ef} = 3,80752, \mu = 6, \rho / c = 0,66667, \\ L_s &= 1,42256, L_q = 0,78797, \\ W_s &= 0,37362, W_q = 0,20695.\end{aligned}$$

Таблиця 2.2 – Вихідні результати для прикладу 2.3

n	P_n	P_n	n	P_n	P_n
0	0,36541	0,36541	3	0,10827	0,87970
1	0,24361	0,60902	4	0,07218	0,95188
2	0,16241	0,77143	5	0,04812	1,00000

2.5.3. Моделі з паралельними сервісами.

У цьому розділі розглядаються три моделі систем масового обслуговування з декількома засобами обслуговування (сервісами), які працюють паралельно. Перші дві моделі є узагальненням моделей, розглянутих у підрозділі 2.5.2, для ситуації декількох паралельно працюючих сервісів. У третій моделі розглядається нескінченна кількість паралельно працюючих сервісів.

Модель $(M / M / c) : (GD / \infty / \infty)$. Ця модель передбачає роботу паралельних засобів обслуговування. Інтенсивність вхідного потоку пакетів дорівнює λ , а інтенсивність обслуговування пакетів — μ для кожного сервісу. Оскільки відсутні обмеження на кількість пакетів в системі, тоді $\lambda_{ef} = \lambda$.

Результатом використання c паралельних сервісів є пропорційне збільшення інтенсивності обслуговування пакетів системою до $n\mu$, якщо $n \leq c$, і до $c\mu$, якщо $n > c$. Отже, в термінах загальної моделі системи обслуговування (розділ 2.4) λ_n и μ_n визначаються наступним чином.

$$\lambda_n = \lambda, \quad n \geq 0,$$

$$\mu_n = \begin{cases} n\mu, & n \leq c, \\ c\mu, & n > c. \end{cases}$$

Відповідно,

$$p_n = \begin{cases} \frac{\lambda^n}{\mu(2\mu)(3\mu)\dots(n\mu)} p_0 = \frac{\lambda^n}{n! \mu^n} p_0 = \frac{\rho^n}{n!} p_0, & n \leq c, \\ \frac{\lambda^n}{\mu(2\mu)\dots(c-1)\mu(c\mu)^{n-c+1}} p_0 = \frac{\lambda^n}{c! c^{n-c} \mu^n} p_0 = \frac{\rho^n}{c! c^{n-c}} p_0, & n > c. \end{cases} \quad (2.17)$$

Значення ймовірності p_0 визначається з рівняння $\sum_{n=0}^{\infty} p_n = 1$. Якщо $\rho = \lambda / \mu$, а $\rho / c < 1$, приходимо до наступної формули для p_0 :

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\rho}{c} \right)^{n-c} \right\}^{-1} = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \left(\frac{1}{1 - \frac{\rho}{c}} \right) \right\}^{-1}, \quad \frac{\rho}{c} < 1. \quad (2.18)$$

Вираз для L_q можливо знайти наступним чином:

$$\begin{aligned} L_q &= \sum_{n=c}^{\infty} (n-c) p_n = \sum_{k=0}^{\infty} k p_{k+c} = \sum_{k=0}^{\infty} k \frac{\rho^{k+c}}{c^k c!} p_0 = \frac{\rho^c \rho}{c! c} \sum_{k=0}^{\infty} k \left(\frac{\rho}{c} \right)^{k-1} p_0 = \\ &= \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} p_0. \end{aligned} \quad (2.19)$$

Оскільки $\lambda_{ef} = \lambda$, то $L_s = L_q + \rho$; значення для W_s и W_q можна знайти, якщо поділити на λ значення L_s і L_q .

Приклад 2.4.

В компанії вихід в Інтернет забезпечується двома провайдерами. Кожен з них встановив маршрутизатори з двома сервісними процесорами (двоканальні системи), і за наявною інформацією пакети даних діляться між провайдерами практично порівну. Це підтверджується тим фактом, що пакети надходять з однією і тією інтенсивністю, яка дорівнює 8 штук у хвилину. Середній час обслуговування (передачі) одного пакета становить 12 секунд. Пакети надходять відповідно до розподілу Пуассона, а час обслуговування клієнтів розподілено по експонентному закону. Компанія зацікавлена в співробітництві тільки з одним провайдером за умови установки одного чотирипроцесорного (чотирьохканальна система) маршрутизатора. Необхідно проаналізувати вигідність нового варіанту мережі для компанії.

З точки зору теорії масового обслуговування процесори представляють собою сервіси. Кожен маршрутизатор може бути представлений моделлю $(M/M/2):(GD:\infty:\infty)$ з $\lambda=8$ пакетів $\mu=60/12=5$ в хвилину пакетів на один процесор в хвилину. Встановлення нового маршрутизатора призведе до моделі $(M/M/4):(GD:\infty:\infty)$ з $\lambda=2 \times 8=16$ пакетів в хвилину і $\mu=5$ пакетів на один процесор в хвилину.

Придатною мірою для порівняння двох моделей обслуговування є середній час очікування пакетом в черзі моменту обслуговування, тобто W_q . У таблиці 2.3 представлені вихідні дані для двох описаних моделей. Результати показують, що час очікування пакетом обслуговування дорівнює 0,356 хвилини (приблизно 21 секунда) для моделі обслуговування з двома маршрутизаторами і 0,149 хвилини (приблизно 9 секунд) для моделі обслуговування в новому варіанті. Значне зменшення (більш ніж на 50%) функціонального показника розглянутої сервісної системи робить наочну доцільність використання нового варіанта.

Таблиця 2.3. Результати розрахунків для прикладу 2.4.

	c	λ	μ	λ_{ef}	ρ_0	L_s	L_q	W_s	W_q
1	2	8	5	8	0,11111	4,44444	2,84444	0,55556	0,35556
2	4	16	5	16	0,02730	5,58573	2,38573	0,34911	0,14911

З наведеного аналізу випливає, що об'єднання систем обслуговування завжди забезпечує більш ефективний режим роботи. Цей висновок залишається справедливим навіть в тому випадку, коли завантаженість всіх сервісів дуже висока.

Модель $(M / M / c) : (GD / N / \infty)$, $c \leq N$.

Ця модель сервісної системи відрізняється від моделі $(M / M / c) : (GD / N / \infty)$ тим, що ємність системи обмежена зверху значенням N (тоді максимальна довжина черги дорівнює $N - c$). Інтенсивності надходження і обслуговування пакетів дорівнюють λ і μ відповідно. Ефективна інтенсивність надходження запитів в систему обслуговування λ_{ef} менша за λ в силу обмеженості ємності системи значенням N .

Параметри λ_n і μ_n загальної моделі обслуговуючої системи в даній моделі визначаються наступним чином:

$$\lambda_n = \begin{cases} \lambda, & 0 \leq n < N, \\ 0, & n \geq N, \end{cases} \quad (2.20)$$

$$\mu_n = \begin{cases} n\mu, & 0 \leq n < c, \\ c\mu, & c \leq n \leq N. \end{cases} \quad (2.21)$$

Підставляючи λ_n і μ_n в загальний вираз (2.3) для p_n з розділу 2.4 та використовуючи позначення $\rho = \lambda / \mu$, отримуємо

$$p_n = \begin{cases} \frac{\rho^n}{n!} p_0, & 0 \leq n < c, \\ \frac{\rho^n}{c! c^{n-c}} p_0, & c \leq n \leq N, \end{cases} \quad (2.22)$$

де

$$p_0 = \begin{cases} \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c \left(1 - \left(\frac{\rho}{c} \right)^{N-c+1} \right)}{c! \left(1 - \frac{\rho}{c} \right)} \right]^{-1}, & \frac{\rho}{c} \neq 1, \\ \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} (N - c + 1) \right]^{-1}, & \frac{\rho}{c} = 1. \end{cases} \quad (2.23)$$

В подальшому обчислюємо L_q для ситуації, коли $\rho / c \neq 1$:

$$\begin{aligned} L_q &= \sum_{n=c}^N (n - c) p_n = \sum_{j=0}^{N-c} j p_{j+c} = \frac{\rho^c \rho}{c! c} \sum_{j=0}^{N-c} j \left(\frac{\rho}{c} \right)^{j-1} p_0 = \frac{\rho^{c+1}}{cc!} \frac{d}{d \left(\frac{\rho}{c} \right)} \sum_{j=0}^{N-c} \left(\frac{\rho}{c} \right)^j = \\ &= \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} \left\{ 1 - \left(\frac{\rho}{c} \right)^{N-c+1} - (N - c + 1) \left(1 - \frac{\rho}{c} \right) \left(\frac{\rho}{c} \right)^{N-c} \right\} p_0. \end{aligned} \quad (2.24)$$

Можливо також показати, що для ситуації, коли $\rho / c = 1$, вираз для L_q має наступний вигляд:

$$L_q = \frac{\rho^c (N - c)(N - c + 1)}{2c!} p_0, \quad \frac{\rho}{c} = 1. \quad (2.25)$$

Для визначення W_q і, відповідно, W_s та L_s , потрібно отримати вираз для λ_{ef} . Оскільки жоден з пакетів не може потрапити до системи після того, як досягнуто ліміту N за її місткістю, тоді

$$\lambda_{втрата} = \lambda p_N, \quad (2.26)$$

$$\lambda_{ef} = \lambda - \lambda_{втрата} = (1 - p_N) \lambda. \quad (2.27)$$

Приклад 2.5.

Нехай в завданні, пов'язаному з об'єднанням систем обслуговування, яка розглядалася в прикладі 2.4, відомо, що черга в маршрутизаторі обмежена і нові вхідні пакети відфільтровуються, коли кількість пакетів, що знаходяться в черзі дорівнює шести. Як така конфігурація системи вплине на її основні характеристики.

Обмеження переліку очікуючих пакетів в черзі до 6 рівнозначно тому, що ємність системи дорівнює $N = 6 + 4 = 10$ (6 пакетів в черзі та 4 пакети на сервісах, які їх обслуговують) пакетів. Отже, завдання зводиться до системи обслуговування моделі $(M / M / 4) : (GD / 10 / \infty)$ з $\lambda = 16$ пакетів в хвилину і $\mu = 5$ пакетів в хвилину. У таблиці 2.4 представлені вихідні дані для цієї моделі.

$$\begin{aligned}\lambda &= 16, \quad \lambda_{ef} = 15,42815, \\ \mu &= 5, \quad \rho / c = 0,8, \\ L_s &= 4,23984, \quad L_q = 1,15421, \\ W_s &= 0,27481, \quad W_q = 0,07481.\end{aligned}$$

Таблиця 2.4 – Вихідні результати для 2.5

n	P_n	P_n	n	P_n	P_n
0	0,03121	0,03121	6	0,08726	0,79393
1	0,09986	0,13106	7	0,06981	0,86374
2	0,15977	0,29084	8	0,05584	0,91958
3	0,17043	0,46126	9	0,04468	0,96426
4	0,13634	0,59760	10	0,03574	1,00000
5	0,10907	0,70667			

Середній час очікування W_q за відсутності обмеження на ємність системи дорівнює 0,149 хвилини (≈ 9 секунд) (див. табл. 2.4), що майже в два рази більше значення 0,075 хвилини ($\approx 4,5$ секунд) — за аналогічний показник наявності обмеження на ємність системи. Це суттєве зменшення функціональної характеристики системи досягнуто за рахунок втрати приблизно 3,6% пакетів.

Модель самообслуговування $(M / M / \infty) : (GD / \infty / \infty)$.

У цій моделі кількість сервісів є необмеженою, так як кожен запит виступає водночас і в ролі сервісу. Прикладом такої системи виступають фонові процеси, що паралельно запускаються, або демони, які, виконавши своє завдання, закриваються.

У розглянутій моделі передбачається, що інтенсивність надходження клієнтів λ є постійною. Інтенсивність обслуговування μ також є постійною. Скориставшись загальною моделлю з розділу 2.4, маємо

$$\begin{aligned}\lambda_n &= \lambda, \quad n = 0, 1, 2, \dots, \\ \mu_n &= n\mu, \quad n = 0, 1, 2, \dots\end{aligned}$$

Таким чином,

$$p_n = \frac{\lambda^n}{n! \mu^n} p_0 = \frac{\rho^n}{n!} p_0, \quad n = 0, 1, 2, \dots \quad (2.28)$$

З рівняння $\sum_{n=0}^{\infty} p_n = 1$ випливає, що

$$p_0 = \frac{1}{1 + \rho + \frac{\rho^2}{2!} + \dots} = \frac{1}{e^\rho} = e^{-\rho}. \quad (2.29)$$

В результаті отримуємо

$$p_n = \frac{e^{-\rho} \rho^n}{n!}, \quad n = 0, 1, 2, \dots \quad (2.30)$$

Ці ймовірності співпадають з ймовірностями розподілу Пуассона з математичним очікуванням $L_s = \rho$. Як і слід було чекати, тут (за принципом самообслуговування) $L_q = W_q = 0$.

Приклад 2.6.

База даних використовує додаткові 1000 мегабайт на годину на диску для програми диспетчера. Оскільки база даних очікує надходження великих транзакцій, фактичний час розширення простору випадковий. База даних зазвичай тримає пам'ять для диспетчера в середньому три 12 вартових періоду, але очищає її у випадковий момент часу, коли закінчиться в ній необхідність. Статистика показує, що близько 25% диспетчерів знижують потребу в пам'яті на 20% в період. Решта 75% диспетчерів підвищують необхідність в пам'яті на 12% в період. Оцінимо середню необхідність в пам'яті для програм диспетчерів бази даних.

Цю ситуацію можливо представити у вигляді моделі $(M / M / \infty) : (GD / \infty / \infty)$, оскільки база даних не повинна чекати черги, щоб запустити або закрити програму диспетчера. В середньому між запусками програм диспетчерів проходить 1 година, що дає значення $\lambda = 12$ запусків в період. Інтенсивність закриття програм диспетчерів дорівнює $\mu = 1/3$ в період.

При вказаних значеннях λ і μ отримуємо

$$L_s = \rho = \frac{\lambda}{\mu} = 36 \text{ програм диспетчерів.}$$

Середній обсяг пам'яті необхідний в один період:

$$(0,25L_s \times 1000)(1 - 0,20) + (0,75L_s \times 1000)(1 + 0,12) = 37\,440 \text{ мегабайт.}$$

2.5.4. Модель $(M / M / R) : (GD / K / K)$ при $R < K$.

Базовим прикладом для цієї моделі є бухгалтерський відділ, що нараховує K комп'ютерів. Кожен раз, коли комп'ютери виходять з ладу, вдаються до послуг одного з працівників служби підтримки, бригада яких складається з R чоловік. Інтенсивність несправностей, віднесена до одного комп'ютера, дорівнює λ несправностей в одиницю часу. Комп'ютерник ремонтує зламані комп'ютери з μ інтенсивністю комп'ютерів в одиницю часу. Передбачається, що моменти часу поломок і час ремонту підкорюються розподілу Пуассона.

Ця модель відрізняється від всіх розглянутих раніше тим, що потужність джерела, що генерує запити, кінцева. Наприклад, в цій моделі джерело може породити кінцеву кількість запитів на ремонт. Це положення стає очевидним, якщо припустити, що всі комп'ютери в бухгалтерії зламані, тоді більше не

надійде жодної запиту на ремонт. По суті, лише працюючі комп'ютери можуть зламатися і, відповідно, генерувати запити на ремонт.

При заданій інтенсивності λ несправностей на один комп'ютер інтенсивність несправностей по всій бухгалтерії пропорційна кількості комп'ютерів в робочому стані. У термінології систем обслуговування наявність n комп'ютерів в системі позначає, що n комп'ютерів зламани. Отже, інтенсивність несправностей по всій бухгалтерії обчислюється так:

$$\lambda_n = (K - n)\lambda, \quad 0 \leq n \leq K. \quad (2.31)$$

В позначеннях загальної системи обслуговування з розділу 2.4 маємо наступне:

$$\lambda_n = \begin{cases} (K - n)\lambda, & 0 \leq n < K, \\ 0, & n \geq K, \end{cases} \quad (2.32)$$

$$\mu_n = \begin{cases} n\mu, & 0 \leq n < R, \\ R\mu, & R \leq n \leq K, \\ 0, & n > K. \end{cases} \quad (2.33)$$

Відтепер із загальної моделі можливо отримати наступні формули:

$$p_n = \begin{cases} C_K^n \rho^n p_0, & 0 \leq n \leq R, \\ C_K^n \frac{n! \rho^n}{R! R^{n-R}} p_0, & R \leq n \leq K, \end{cases} \quad (2.34)$$

$$p_0 = \left\{ \sum_{n=0}^R C_K^n \rho^n + \sum_{n=R+1}^K C_K^n \frac{n! \rho^n}{R! R^{n-R}} \right\}^{-1}. \quad (2.35)$$

В цій моделі важко отримати в замкненій формі вираз для L_s або L_q , та, відповідно, вони повинні обчислюватися у відповідності з їх визначенням:

$$L_s = \sum_{n=0}^K n p_n. \quad (2.36)$$

Значення λ_{ef} можливо визначити в наступній формі:

$$\lambda_{ef} = M \{ \lambda (K - n) \} = \lambda (K - L_s). \quad (2.37)$$

Використовуючи формули (2.4) – (2.8), можливо обчислити полишені функціональні показники W_s , W_q і L_q .

Приклад 2.7.

Відділ бухгалтерії налічує 22 комп'ютери. Відомо, що кожен комп'ютер виходить з ладу в середньому один раз на дві години. На його ремонт йде в середньому 12 хвилин. Як час між несправностями комп'ютерів, так і час, необхідний для ремонту, є експоненційно розподіленими випадковими величинами. Компанія зацікавлена у визначенні мінімальної кількості комп'ютерів, необхідних для забезпечення «плавної» роботи бухгалтерії.

Описану ситуацію можливо проаналізувати, досліджуючи продуктивність комп'ютерів як функцію кількості. Така міра продуктивності може бути визначена в наступному вигляді.

$$\begin{aligned} \left(\frac{\text{Продуктивність}}{\text{комп'ютерів}} \right) &= \frac{\text{Кількість комп'ютерів} - \text{зламани комп'ютери}}{\text{Кількість комп'ютерів}} \times 100 = \\ &= \frac{22 - L_s}{22} \times 100. \end{aligned}$$

У табл. 2.5 представлені отримані порівняльні характеристики сервісної системи для $R=1,2,3,4$ при $\lambda=0,5$ несправностей на годину на один комп'ютер і $\mu=5$ ремонтів на годину. Відповідні значення продуктивності наведені в наступній таблиці.

Таблиця 2.5– Параметри роботи при різних кількостях комп'ютерів

Кількість комп'ютерів, R	1	2	3	4
Продуктивність комп'ютерів (%)	45,44	80,15	88,79	90,45
Зріст продуктивності (%)	—	34,71	8,64	1,66

Наведені результати показують, що неприйнятною є наявність лише одного комп'ютера, оскільки в цьому випадку буде дуже низька продуктивність

(45,44%). Якщо число комп'ютерів збільшується до двох, то продуктивність бухгалтерів зростає на 34,71% і досягає 80,15%. При роботі трьох ремонтників продуктивність бухгалтерів зростає приблизно на 8,64% і досягає значення 88,79%, в той час як при наявності чотирьох комп'ютерів продуктивність зростає лише на 1,66% і досягає значення 90,45%.

Судячи за цими результатами, використання двох ремонтників може бути виправданим. Ситуація з трьома ремонтниками є не такою переконливою, оскільки продуктивність при цьому зростає лише на 8,64%. Можливо, що порівняння в грошовому еквіваленті утримання третього ремонтника і прибутку, зумовленої зростанням продуктивності бухгалтерії, на 8,64% можна використовувати для вирішення цього питання. Що стосується прийому на роботу четвертого ремонтника, то надто мале зростання продуктивності бухгалтерії на 1,66%, який при цьому досягається, не виправдовує цього кроку.

Таблиця 2.6 – Порівняльні характеристики систем з прикладу 2.7.

N	c	λ	μ	λ_{ef}	p_0	L_s	L_q	W_s	W_q
1	1	0,5	5	4,99798	0,00040	12,00404	11,00444	2,40178	2,20178
2	2	0,5	5	8,81616	0,05638	4,36768	2,60447	0,49542	0,29542
3	3	0,5	5	9,76703	0,10779	2,46593	0,51257	0,25247	0,05248
4	4	0,5	5	9,94995	0,11993	2,10010	0,11015	0,21107	0,01107

2.6. Модель $(M / G / 1) : (GD / \infty / \infty)$.

Аналіз моделей масового обслуговування, в яких вхідні і вихідні потоки не підкорюються пуассонівському розподілу, досить складний. Взагалі, в таких випадках в якості альтернативного апарату для аналізу моделей обслуговування доцільно використовувати методи імітаційного моделювання.

У цьому розділі розглядається один з небагатьох варіантів системи масового обслуговування, що не підкорюється пуассонівському розподілу, для якого можуть бути отримані аналітичні результати. Мова йде про той випадок, коли час обслуговування t має довільний розподіл з математичним очікуванням $M\{t\}$ і дисперсією $D\{t\}$. Для такої моделі відомі аналітичні формули для основних функціональних характеристик сервісної системи, таких

як, L_s , L_q , W_s и W_q . Але в силу аналітичних труднощів неможливо отримати в замкнутій формі вираз для ймовірностей p_n .

Нехай λ – інтенсивність надходження запитів в систему обслуговування з одним сервісом. При заданих значеннях $M\{t\}$ і $D\{t\}$ для часу обслуговування і умови $\lambda M\{t\} < 1$ з використанням складного аналізу, пов'язаного із застосуванням апарату ланцюгів Маркова, можна показати, що

$$L_s = \lambda M\{t\} + \frac{\lambda^2 (M^2\{t\} + D\{t\})}{2(1 - \lambda M\{t\})}, \quad \lambda M\{t\} < 1. \quad (2.38)$$

Ймовірність, що сервіс буде не завантаженим, обчислюється так

$$p_0 = 1 - \lambda M\{t\} = 1 - \rho. \quad (2.39)$$

Оскільки $\lambda_{ef} = \lambda$, тоді залишені функціональні характеристики сервісної системи (L_q , W_s и W_q) можливо отримати з формули для L_s , як це зроблено у розділі 2.5.1.

Приклад 2.8.

Нехай у завданні про маршрутизатор з прикладу 2.1 зроблено додаткове припущення, що встановлено нове мережне обладнання для синхронної передачі даних, тому час обслуговування всіх пакетів є постійним і дорівнює 10 хв. Необхідно визначити, як нове обладнання впливає на функціональні характеристики вузла мережі.

На підставі даних прикладу 2.1 маємо, що $\lambda_{ef} = \lambda = 4$ пакета в хвилину. Час обслуговування є постійним, отже $M\{t\} = 10 / 60 = 1 / 6$ хвилини і $D\{t\} = 0$.

Відповідно,

$$L_s = 4 \left(\frac{1}{6} \right) + \frac{4^2 \left[\left(\frac{1}{6} \right)^2 + 0 \right]}{2 \left(1 - \frac{4}{6} \right)} = 1,333 \text{ пакета,}$$

$$L_q = 1,333 - \left(\frac{4}{6} \right) = 0,667 \text{ пакета,}$$

$$W_s = \frac{1,333}{4} = 0,333 \text{ хвилини},$$

$$W_q = \frac{0,667}{4} = 0,167 \text{ хвилини.}$$

Цікаво відзначити, що, незважаючи на те, що інтенсивності, як надходження пакетів, так і обслуговування в розглянутій моделі саме такі, як і в пуассонівському випадку з прикладу 2.1 ($\lambda = 4$ пакета в хвилину і $\mu = 1 / M\{t\} = 6$ пакета в хвилину), середній час очікування в даному випадку менше, оскільки час обслуговування є постійним, про що свідчить таблиця 2.7.

Таблиця 2.7 – Порівняння W_s та W_q для систем с різними параметрами.

	$(M / M / 1) : (GD / \infty / \infty)$	$(M / D / 1) : (GD / \infty / \infty)$
W_s (хвилини)	0,5	0,333
W_q (хвилини)	0,333	0,167

У цих результатах є сенс, оскільки постійний час обслуговування має на увазі велику визначеність у функціонуванні системи.

2.7. Моделі прийняття рішень в теорії масового обслуговування.

Рівень обслуговування в системі є функцією інтенсивності обслуговування μ і кількості s паралельно працюючих сервісів. У цьому розділі розглядаються дві моделі прийняття рішень для визначення відповідних рівнів обслуговування для систем масового обслуговування:

- 1) модель з вартісними характеристиками.
- 2) модель бажаного рівня обслуговування.

В обох моделях більш високий рівень обслуговування має на увазі зменшення часу очікування в системі. У цих моделях для пошуку рівноваги між конфліктуючими факторами (рівнем обслуговування і часом очікування в системі) використовуються функціональні показники сервісної системи, які отримані раніше для різних моделей.

2.7.1. Модель з вартісними характеристиками

Моделі з вартісними характеристиками прагнуть зрівноважити два конфліктуючих вартісних показника.

1. Витрати на обслуговування.

2. Втрати, зумовлені затримками в наданні послуг (час очікування клієнта).

Ці два види витрат конфліктують між собою, оскільки збільшення одного з них автоматично веде до зменшення іншого і навпаки (рис.2.1).

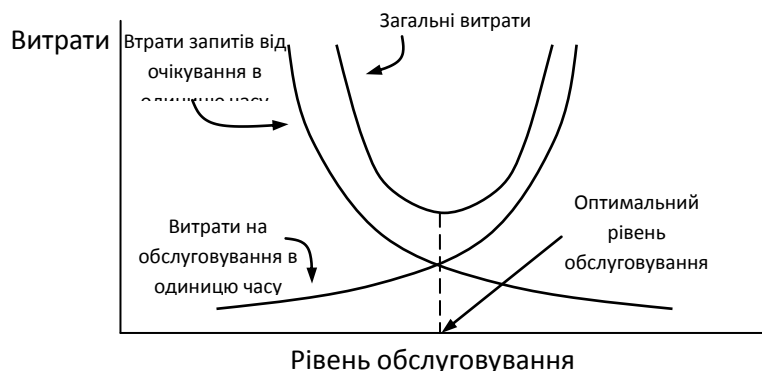


Рис. 2.1. Вартісна модель системи обслуговування

Нехай рівень обслуговування представляє змінна x , яка дорівнює μ або c . Тоді модель з вартісними характеристиками можна представити в наступному вигляді:

$$COC(x) = CTC(x) + CTO(x), \quad (2.40)$$

де,

COC – середня загальна вартість в одиницю часу,

CTC – середня вартість обслуговування в одиницю часу,

CTO – середня вартість очікування в одиницю часу.

Найпростішим видом функцій CTC та CTO є лінійні функції:

$$CTC(x) = C_1 x, \quad (2.41)$$

$$CTO(x) = C_2 L_s, \quad (2.42)$$

де

C_1 – питома вартість на одиницю x в одиницю часу,

C_2 – величина штрафу за очікування в одиницю часу на (очікуючого) клієнта.

Наступні два приклади ілюструють використання вартісної моделі. У першому прикладі передбачається, що x дорівнює інтенсивності обслуговування, у другому — кількості паралельних сервісів c .

Приклад 2.9.

Видавнича фірма купує високошвидкісний копіювальний апарат для комерційних цілей. Продавці запропонували чотири моделі копіювальних апаратів, характеристики яких наведені в таблиці нижче.

Таблиця 2.8 – Характеристики копіювальних апаратів

Модель копіювального апарату	Експлуатаційні витрати (грн/год)	Швидкість друку (стор/хв)
1	15	30
2	20	36
3	24	50
4	27	66

Замовлення надходять на фірму відповідно до пуассонівського розподілу з математичним очікуванням з інтенсивністю чотири роботи протягом 24-годинного дня. Обсяг роботи є випадковою величиною, але в середньому становить приблизно 10000 сторінок. Договори з клієнтами передбачають штраф у сумі 80 грн. (за одну роботу) за затримку виконання замовлення на один день. Необхідно визначити, який копіювальний апарат слід купити фірмі.

Нехай індекс i представляє собою номер моделі копіювального апарату ($i=1,2,3,4$). Загальна очікувана вартість обслуговування в день, пов'язана з використанням копіювального апарату i , представлена в наступному вигляді:

$$COC_i = CTC_i + CTO_i = C_{1i} \times 24 + C_{2i} \times L_{si} = 24C_{1i} + 80L_{si}, \quad i=1,2,3,4.$$

Значення величин наведені в постановці завдання. Визначимо значення L_{si} , вважаючи, що з практичної точки зору копіювальний апарат може розглядатися як модель $(M/M/1): (GD/\infty/\infty)$ сервісної системи. При цьому інтенсивність надходження запитів $\lambda = 4$ роботи в день, а інтенсивності

μ_i обслуговування, відповідні моделі i -го копіювального апарату, наведені в наступній таблиці.

Таблиця 2.9 – Інтенсивності обслуговування μ_i

Модель i	Інтенсивність обслуговування μ_i (роботи / доба)
1	4,32
2	5,18
3	7,20
4	9,50

Проілюструємо визначення інтенсивності обслуговування на прикладі моделі 1.

$$\text{Середній час виконання роботи} = \frac{1000}{30} \times \frac{1}{60} = 5,56 \text{ год}$$

Відповідно,

$$\mu_1 = \frac{24}{5,56} = 4,32 \text{ роботи на добу.}$$

Значення L_{si} наведені в наступній таблиці.

Таблиця 2.10 – Значення L_{si}

Модель i	λ_i (роботи/доба)	μ_i (роботи/доба)	L_{si} (роботи)
1	4	4,32	12,50
2	4	5,18	3,39
3	4	7,20	1,25
4	4	9,50	0,73

Отримані наступні вартісні характеристики (в грн) для чотирьох розглянутих моделей копіювальних апаратів.

Таблиця 2.11 – Вартісні характеристики розглянутих моделей

Модель i	CTC_i	CTO_i	COC_i
1	360	1000	1360

2	480	271,20	751,20
3	576	100	676
4	648	58,40	706,40

Відповідно, для третьої моделі копіювального апарату вартість мінімальна.

Приклад 2.10.

Провайдеру, де працюють кілька мережних майстрів, надходять запити від користувачів на проблеми роботи Інтернету відповідно до розподілу Пуассона із середнім значенням 17,5 запитів на годину. Кожен мережний майстер може виконати в середньому 10 запитів на годину. Вартість найму додаткового мережного майстра становить 12 грн. в годину. Втрати, пов'язані з очікуванням роботи Інтернету клієнтами, оцінюються приблизно в 50 грн. в годину. Необхідно визначити число мережних майстрів для даної системи обслуговування.

Описана ситуація відповідає моделі $(M/M/c)$, в якій потрібно визначити оптимальне значення c . Отже, $x=c$, і відповідна вартісна модель має вигляд

$$COC(c) = C_{1c} + C_2 L_s(c) = 12c + 50L_s(c).$$

Зауважимо, що $L_s(c)$ є функцією кількості мережних майстрів.

Використовуємо модель $(M/M/c):(GD/\infty/\infty)$ з $\lambda=17,5$ запитів на годину і $\mu=10$ запитів на годину. В цьому відношенні модель досягне стійкого стану лише за умови, що $c > \lambda/\mu$, тобто в розглянутому прикладі c має, принаймні, дорівнювати 2. Наведена нижче таблиця містить результати обчислень для визначення оптимального значення c . Значення $L_s(c)$ показані в таблиці.

Таблиця 2.12 – Результати обчислень для прикладу 2.10.

c	$L_s(c)$ (запиту)	$COC(c)$ (грн.)
2	7,467	397,35

3	2,217	142,35
4	1,842	140,10
5	1,769	148,45
6	1,754	159,70

2.7.2. Модель кращого рівня обслуговування.

Життєздатність моделі сервісної системи з вартісними характеристиками залежить від того, наскільки добре можна оцінити параметри вартості. У загальному випадку оцінити ці параметри досить складно, особливо якщо вартість пов'язана з очікуванням пакету. У моделях з переважним рівнем обслуговування робиться спроба обійти цю проблему, оперуючи безпосередньо функціональними показниками сервісної системи. Ідея полягає у визначенні прийняттого інтервалу зміни для рівня обслуговування (параметри μ або c) шляхом пошуку розумних меж для конкуруючих економічних показників, які характеризують процес обслуговування. Ці межі представляють собою рівні кращого обслуговування, яких прагне досягти особа, яка приймає управлінське рішення.

Проілюструємо застосування цієї процедури для моделі системи обслуговування з декількома сервісами, в якій необхідно визначити прийнятну кількість сервісів c^* . Для цього розглянемо два конкуруючих економічних показника процесу обслуговування.

1. Середній час очікування в системі W_s .

2. Відсоток простою сервісів X .

Значення W_s можна обчислити за відомими формулами. Відсоток простою засобів обслуговування можна обчислити таким чином.

$$X = \frac{c - \bar{c}}{c} \times 100 = \frac{c - (L_s - L_q)}{c} \times 100 = \left(1 - \frac{\lambda_{ef}}{c\mu} \right) \times 100. \quad (2.43)$$

Завдання зводиться до визначення такої кількості сервісів c^* , що

$$W_s \leq \alpha \text{ и } X \leq \beta,$$

де α і β – рівні переважного обслуговування, визначені особою, яка приймає рішення. Наприклад, можна поставити умову, що $\alpha = 3$ хв. і $\beta = 10\%$.

Завдання можна вирішити, побудувавши графіки W_s і X як функції кількості сервісів c (рис. 2.5). Відзначаючи на графіках значення α і β , визначається прийнятний інтервал зміни для рівня обслуговування c^* . Якщо обидві згадані вище умови не можна задовольнити одночасно, необхідно послабити один або обидва рівня переваги, поки не буде отримано прийнятний інтервал зміни кількості сервісів.

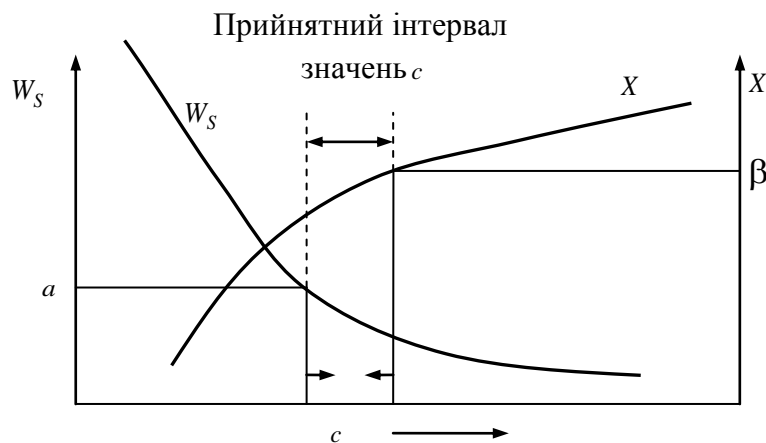


Рис. 2.2. Прийнятний інтервал зміни для рівня обслуговування

Приклад 2.11.

Припустимо, що в прикладі 2.10 необхідно визначити таку кількість ремонтників у провайдера, щоб середній час очікування ремонту був менший за 5 хв. Водночас потрібно, щоб відсоток часу, протягом якого персонал провайдера залишається вільним, не перевищував 20%.

Зауважимо, що й без обчислень очевидним є той факт, що рівень переваги в 5 хв. для часу очікування ремонту (тобто тут $W_s \leq 5$) є недосяжним, оскільки з вхідних даних задачі випливає, що середній час обслуговування дорівнює 6 хв. Наведена нижче таблиця містить значення W_s і X як функцій від c , і підтверджує це зауваження. Дійсно при $c \geq 5$ маємо $W_s = 6$ хв. Таблиця 2.13 також показує, що подальше збільшення числа ремонтників може лише підвищити відсоток їх простою X .

Таблиця 2.13 – Залежність W_s и X от c

c	2	3	4	5	6	7	8
-----	---	---	---	---	---	---	---

W_s (хвилини)	25,6	7,6	6,3	6,1	6,0	6,0	6,0
X (%)	12,5	41,7	56,3	65,0	70,8	75,0	78,0

У цій ситуації фактично нічого не можливо зробити, так як завдання не можна вирішити шляхом збільшення кількості комп'ютерників у провайдера. Необхідно або зменшити час обслуговування в системі, або визнати, що в реальній ситуації, якій відповідає математична модель, інтенсивність запитів на інструменти надзвичайно висока ($\lambda = 17,5$ запитів на годину). Досить ймовірно, що саме на останній факт має бути звернено увагу. Наприклад, можна провести дослідження причини такого високого попиту на заміну інструменту.

Висновки.

Таким чином, у даному розділі отримано математичні співвідношення для обчислення основних стаціонарних характеристик варіацій систем масового обслуговування, функціонуючих в дискретному часі та з дискретною множиною станів.

Отримані співвідношення можуть бути використані для написання програм для імітаційного та математичного моделювання вузлів інфокомунікаційних мереж на вхід яких поступає однорідний потік пакетів.

Розрахунки показників якості обслуговування у вузлах мереж для ситуації із неоднорідним вхідним потоком буде розглянуто у наступних розділах.

Контрольні питання.

1. Що таке однорідний трафік?
2. Поясніть маркування моделі $(M/M/1):(GD/\infty/\infty)$. Вкажіть формули розрахунку середньої довжини черги, середньої кількості заявок у системі, середню зайнятість засобів обслуговування, середній час обслуговування та очікування у черзі.
3. Поясніть маркування моделі $(M/M/1):(FCFS/\infty/\infty)$. Надайте формули розрахунку часу очікування в цій моделі.
4. Поясніть маркування моделі $(M/M/1):(GD/N/\infty)$. Надайте формули розрахунку значень L_s , $\lambda_{эфф}$, W_s , W_q та L_q .
5. Надайте приклади моделей з паралельними сервісами.
6. Надайте розрахунки загальних характеристик моделі $(M/M/c):(GD/\infty/\infty)$.
7. Надайте розрахунки загальних характеристик моделі $(M/M/c):(GD/N/\infty)$, $c \leq N$.

8. Надайте розрахунки загальних характеристик моделі самообслуговування $(M/M/\infty):(GD/\infty/\infty)$.
9. Надайте розрахунки загальних характеристик моделі $(M/M/R):(GD/K/K)$ при $R < K$.
10. Надайте розрахунки загальних характеристик моделі $(M/G/1):(GD/\infty/\infty)$.

3 Моделі вузлів мереж з неоднорідним трафіком з пріоритетами.

3.1. Інфокомунікаційні мережі, як системи масового обслуговування.

Телекомунікаційні мережі створені для вирішення сукупності завдань обробки потоків запитів на обслуговування. Програми P_1, \dots, P_g вирішення завдань обслуговування зберігаються в постійній або оперативній пам'яті інтелектуального вузла. Обсяг, структура та зміст задач з різних потоків — різні. Тому є різними і програми, які їх оброблюють. Запити, що надходять на вхід мережі, ініціюються в порядку, визначеному процесами, що відбуваються в середовищі передачі телекомунікаційної мережі та в самому вузлі. Вони генеруються в об'єктах джерел мережі і надходять у вузол–приймач періодично або в будь-які, випадкові моменти часу. При цьому за короткий період часу може надійти декілька запитів z_a, \dots, z_w , для обслуговування яких повинні бути виконані відповідні програми P_a, \dots, P_w обслуговування вузлів. При наявності одного процесора у вузлі обслуговування ці програми можуть бути виконані тільки послідовно, у зв'язку з чим виникають черги запитів на обслуговування. При наявності декількох процесорів (наприклад, в комутаторах) черги на обслуговування (запит на передачу пакета в порт призначення може бути поставлений в чергу) формуються у зв'язку із зайнятістю порта призначення [8].

Обробка запитів у вузлах ТКС організується по схемі, що показана на рис. 3.1. Запити z_1, \dots, z_m надходять в пристрій переривання, що є частиною вузла мережі. При надходженні запиту z_i пристрій переривання ініціює у вузлі операцію переривання, в результаті виконання якої вузол перемикається на виконання програми прийому та постановки запитів в чергу (програма "Диспетчер 1").

"Диспетчер 1" визначає тип поданого запиту та встановлює запит в відповідну чергу O_j на обслуговування. Черга у фізичному відношенні складається з сукупності комірок оперативної пам'яті, в яких розміщені коди надісланих заяв. В загальному випадку кожна з черг містить запити, що очікують обслуговування. Нехай у будь-який момент часу вузол може виконати лише одну програму обслуговування. Процес вибору запиту з безлічі запитів, що очікують обслуговування, називається диспетчеризацією.

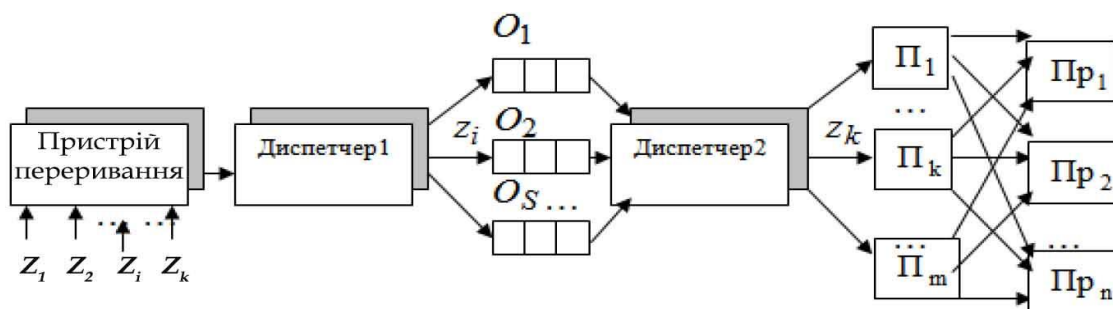


Рис. 3.1. Обробка запитів в узлі ТКС

Процедура диспетчеризації реалізується програмою, яка називається "Диспетчер 2", яка аналізує стан черг O_1, \dots, O_N , обирає запит z_k , що має переважне право на обслуговування, та ініціює відповідну прикладну програму Π_k . Вважається, що в момент закінчення роботи програми виконаний запит залишає систему. Після закінчення програми Π_k керування знову надається "Диспетчеру 2", яка вибирає черговий запит та ініціює відповідну прикладну програму. Якщо черги відсутні, "Диспетчер 2" перемикає процесор у стан очікування.

Відзначимо важливі особливості процесу функціонування вузла ТКС. Потoki запитів, що надходять в систему, є випадковими. Точно так само випадковими для кожного з запитів є тривалість очікування початку обслуговування та, власне, тривалість самого обслуговування. Таким чином, сам процес функціонування вузла ТКС носить стохастичний характер, що дозволяє розглядати такі системи як системи масового обслуговування (СМО) [6].

Аналіз СМО може бути виконаний аналітично лише при деяких досить жорстких припущеннях щодо характеру вхідного потоку та розподілу тривалості обслуговування.

Якщо вхідний потік найпростіший, тривалість обслуговування експонентна, аналіз може бути проведений з використанням теорії марківських процесів [4]. При цьому ця технологія аналізу є найпростішою, якщо потоки запитів однорідні й, отже, їх суперпозицію можливо розглядати як єдиний однорідний вхідний потік [6]. Тоді такі завдання стандартно вирішуються із застосуванням формул Ерланга [2].

3.2. Модель інтерфейсу вузла мережі з неоднорідним вхідним потоком пакетів без переваг.

Розглянемо технологію аналізу вузла ТКС, використовуючи марківські моделі їх функціонування [2].

Нехай на вхід вузла надходить суперпозиція двох пуассонівських потоків з інтенсивностями λ_1 і λ_2 . Уявімо, що інтенсивність обслуговування запитів цих потоків дорівнює відповідно μ_1 і μ_2 . Будемо вважати надалі, що довжини черг для потоків запитів першого та другого типів обмежені значеннями n_1 і n_2 . Введемо множину можливих станів вузла. Нехай E_{ik} — стан вузла, який відповідає ситуації, коли в системі знаходяться i запитів першого типу та k запитів другого типу. Граф станів та переходів такої системи наведено на малюнку 3.2.

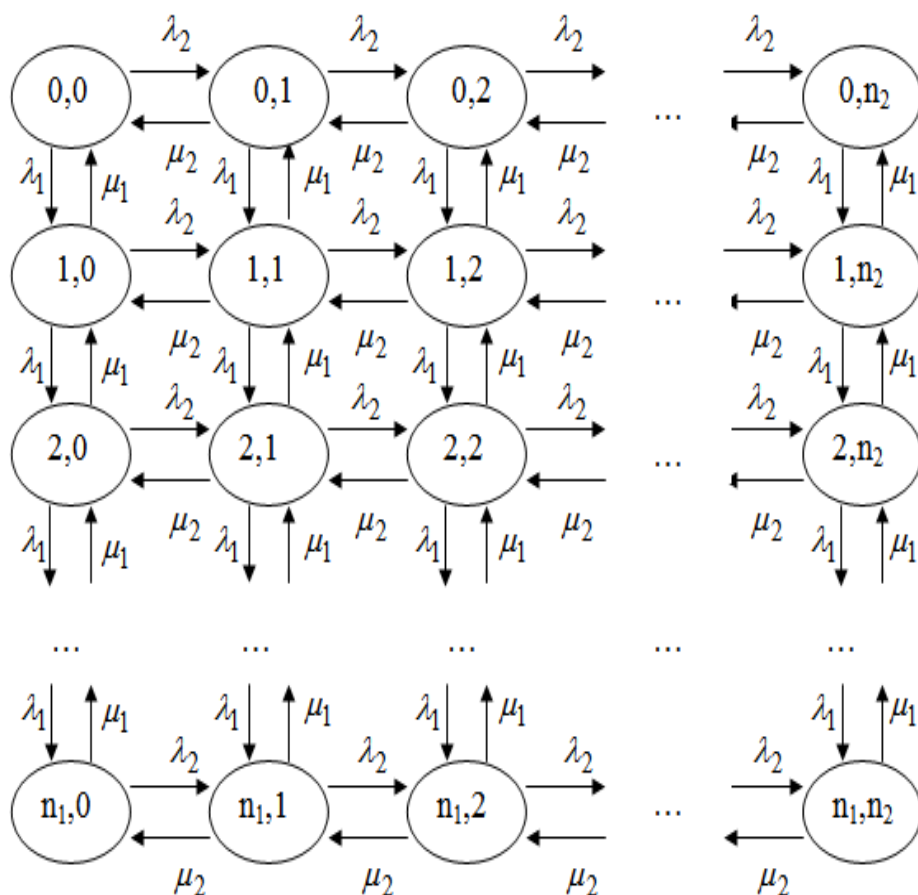


Рисунок 3.2. Граф станів та переходів системи.

Для вирішення завдання використовуємо технологію фазового укрупнення станів. Будемо говорити, що безліч станів $\{E_{i0}, E_{i1}, E_{i2}, \dots, E_{in_2}\}$

належать шару i . Введемо укрупнений стан $E_{\Gamma i} = \{ E_{i0}, E_{i1}, E_{i2}, \dots, E_{in_2} \}$. Тоді в термінах укрупнених станів граф станів переходів прийме такий вигляд (рис. 3.3):

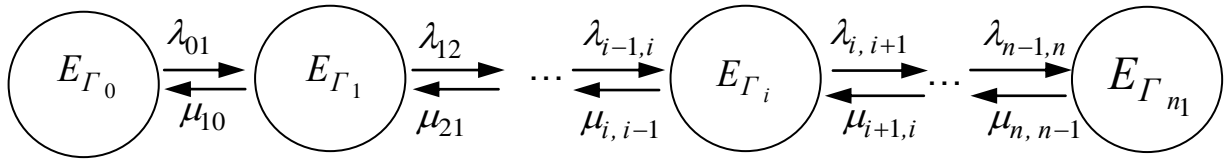


Рис. 3.3. Граф станів та переходів укрупненої системи.

Розрахуємо інтенсивності переходів для графа укрупнених станів. Інтенсивність $\lambda_{i,i+1}$ переходу з $E_{\Gamma i}$ в $E_{\Gamma,i+1}$ дорівнює

$$\lambda_{i,i+1} = \sum_{k=0}^{n_2} \lambda_l \cdot \hat{P}_{ik} = \lambda_l \sum_{k=0}^{n_2} \hat{P}_{ik}, i=0,1,\dots,n_l-1$$

де \hat{P}_{ik} – умовна ймовірність перебування системи в k -ому стані i -го шару, за умови знаходження в цьому шарі.

Оскільки $\sum_{k=0}^{n_2} \hat{P}_{ik} = 1$, то $\lambda_{i,i+1} = \lambda_l$, $i=0,1,\dots,n_l-1$

Саме так знайдемо інтенсивність переходу зі стану $E_{\Gamma,i+1}$ в $E_{\Gamma i}$.

Маємо

$$\mu_{i+1,i} = \sum_{k=0}^{n_2} \mu_l \cdot \hat{P}_{i+1,k} = \mu_l \sum_{k=0}^{n_2} \hat{P}_{i+1,k} = \mu_l, i=0,1,\dots,n_l-1.$$

Знайдемо тепер розподіл ймовірностей перебування системи в кожному з шарів, відповідних укрупненим станів. З урахуванням отриманих співвідношень для інтенсивностей переходів граф станів і переходів набуває вигляду, представлений на малюнку 3.4.

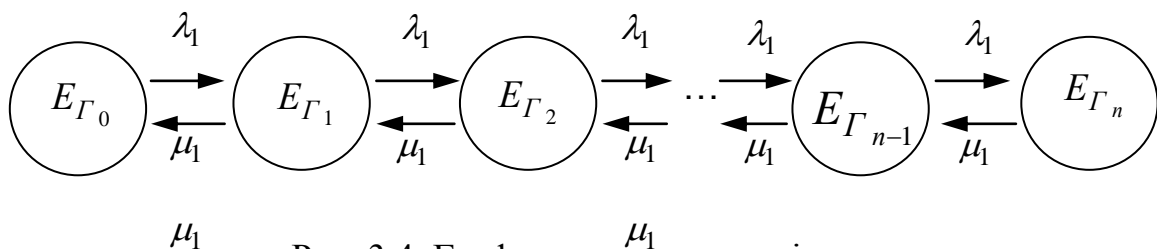


Рис. 3.4. Граф укрупнених станів

Запишемо відповідну систему рівнянь для ймовірностей укрупнених станів.

$$\left\{ \begin{array}{l} \mu_1 P - \lambda_1 P_0 = 0, \\ \lambda_1 P_0 + \mu_1 P_2 (\lambda_1 + \mu_1) P_1 = 0, \\ \dots\dots\dots \\ \lambda_1 P_{i-1} + \mu_1 P_{i+1} - (\lambda_1 + \mu_1) P_i = 0, \\ \dots\dots\dots \\ \lambda_1 P_{n_1-1} - \mu_1 P_{n_1} = 0 \end{array} \right. \quad (3.1)$$

Введемо $z_i = \lambda_1 P_{i-1} - \mu_1 P_i$ [240]. Тоді система рівнянь (3.1) прийме вигляд:

$$\left\{ \begin{array}{l} z_1 = 0, \\ z_1 - z_2 = 0, \\ \dots\dots\dots \\ z_i - z_{i+1} = 0, \\ z_{n_1} = 0. \end{array} \right.$$

Звідси $z_1 = z_2 = \dots = z_{n_1} = 0$, тобто

$$\lambda_1 P_{i-1} + \mu_1 P_i = 0, i = 1, 2, \dots, n_1$$

При цьому

$$P_i = \frac{\lambda_1}{\mu_1} P_{i-1}, i = 1, 2, \dots, n_1 .$$

Тоді

$$\begin{aligned}
P_1 &= \frac{\lambda_l}{\mu_l} P_0, \\
P_2 &= \frac{\lambda_l}{\mu_l} P_1 = \left(\frac{\lambda_l}{\mu_l} \right)^2 P_0, \\
&\dots\dots\dots \\
P_i &= \left(\frac{\lambda_l}{\mu_l} \right)^i P_0, \\
&\dots\dots\dots \\
P_n &= \left(\frac{\lambda_l}{\mu_l} \right)^n P_0.
\end{aligned}$$

Значення P_0 знаходиться з умови нормування

$$\sum_{i=0}^n P_i = P_0 \sum_{i=0}^{n_i} \left(\frac{\lambda_l}{\mu_l} \right)^i = 1$$

Звідси

$$P_0 = \frac{1}{\sum_{i=0}^{n_i} \left(\frac{\lambda_l}{\mu_l} \right)^i}, i = 1, 2, \dots, n_i.$$

Тоді

$$P_i = \frac{\left(\frac{\lambda_l}{\mu_l} \right)^i}{\sum_{i=0}^{n_i} \left(\frac{\lambda_l}{\mu_l} \right)^i}, i = 1, 2, \dots, n_l.$$

Введемо $\rho_l = \frac{\lambda_l}{\mu_l}$. При цьому

$$P_i = \frac{\rho_1^i}{\sum_{i=0}^{n_i} \rho_1^i}, i = 1, 2, \dots, n_i. \quad (3.2)$$

Оскільки $\rho_1 < 1$, тоді послідовність $1, \rho_1, \rho_1^2, \dots, \rho_1^{n_i}$ утворює спадаючу геометричну прогресію з першим членом, який дорівнює 1, і знаменником ρ ,
Тому

$$\sum_{i=0}^{n_i} \rho_1^i = \frac{1 - \rho_1^{n_i+1}}{1 - \rho_1}.$$

Тоді розподіл ймовірностей укрупнених станів має вигляд

$$P_i = \frac{\rho_1^i (1 - \rho_1)}{1 - \rho_1^{n_i+1}}, i = 0, 1, 2, \dots, n_i. \quad (3.3)$$

Знайдемо тепер умовні розподіли ймовірностей станів системи для кожного з шарів. Ці розподіли, вочевидь, однакові, оскільки граф станів і переходів для кожного з них має вигляд, показаний на рис. 3.5.

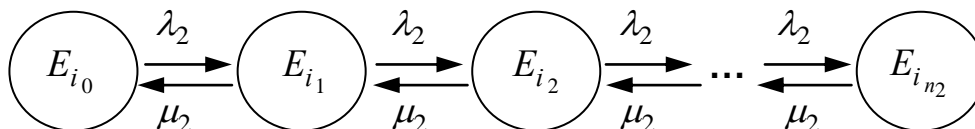


Рис. 3.5. Граф станів і переходів в i -ому шарі

Технологія аналізу цього графу станів абсолютно аналогічна описаній вище. Умовні ймовірності дорівнюють:

$$P_{i,k} = \frac{\rho_2^k (1 - \rho_2)}{1 - \rho_2^{n_i+1}}, k = 0, 1, 2, \dots, n_i. \quad (3.4)$$

де $\rho_2 = \frac{\lambda_2}{\mu_2}$.

Тоді безумовна ймовірність перебування системи в i -ому стані k -го шару визначається співвідношенням

$$P_{i,k} = P_{I^k} \cdot P_i = \frac{\rho_1^i \cdot \rho_2^k (1 - \rho_1)(1 - \rho_2)}{(1 - \rho_1^{n_1+1})(1 - \rho_2^{n_2+1})}, i = 0, 1, \dots, n; k = 0, 1, \dots, n_2 \quad (3.5)$$

Якщо довжина черги для запитів обох типів необмежена, тоді співвідношення (3.5) спрощується до виду

$$P_{i,k} = \rho_1^i \cdot \rho_2^k (1 - \rho_1)(1 - \rho_2). \quad (3.6)$$

Розрахуємо середні значення \bar{n}_1 , \bar{n}_2 числа запитів першого і другого типу, що стоять в черзі. Зрозуміло, що

$$\begin{aligned} \bar{n}_1 &= \sum_{i=1}^{\infty} i \sum_{k=1}^{\infty} P_{i,k} = \sum_{i=1}^{\infty} i \sum_{k=1}^{\infty} \rho_1^i \rho_2^k (1 - \rho_1)(1 - \rho_2) = \\ &= (1 - \rho_1)(1 - \rho_2) \sum_{i=1}^{\infty} i \rho_1^i \sum_{k=1}^{\infty} \rho_2^k = (1 - \rho_1) \sum_{i=1}^{\infty} i \rho_1^i = (1 - \rho_1) \rho_1 \sum_{i=1}^{\infty} i \rho_1^{i-1} \\ &= (1 - \rho_1) \rho_1 \sum_{i=1}^{\infty} \frac{d(\rho_1^i)}{d\rho_1} = (1 - \rho_1) \rho_1 \frac{d}{d\rho_1} \sum_{i=1}^{\infty} \rho_1^i = \\ &= (1 - \rho_1) \rho_1 \frac{d}{d\rho_1} \left(\frac{1}{1 - \rho_1} \right) = \frac{(1 - \rho_1) \rho_1}{(1 - \rho_1)^2} = \frac{\rho_1}{1 - \rho_1} \end{aligned} \quad (3.7)$$

Аналогічно цьому

$$\begin{aligned} \bar{n}_2 &= \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} k P_{i,k} = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} k \rho_1^i \rho_2^k (1 - \rho_1)(1 - \rho_2) = \\ &= (1 - \rho_1) \sum_{k=0}^{\infty} k \rho_2^k = \frac{\rho_2}{1 - \rho_2} \end{aligned} \quad (3.8)$$

Тепер легко розрахувати середню тривалість очікування запитів в черзі до початку обслуговування. Нехай черговий запит надходить в систему в момент,

коли в черзі знаходяться i запитів першого типу і k запитів другого. При цьому середній час очікування початку обслуговування для вхідного запиту, дорівнює

$$\bar{\tau}_{ik} = \frac{i}{\mu_1} + \frac{k}{\mu_2}$$

Усреднюючи це співвідношення по всім i і k , отримаємо

$$\bar{\tau} = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \left(\frac{i}{\mu_1} + \frac{k}{\mu_2} \right) P_{i,k} = \sum_{i=0}^{\infty} \frac{1}{\mu_1} \sum_{k=0}^{\infty} P_{i,k} + \sum_{k=1}^{\infty} \frac{1}{\mu_2} \sum_{i=0}^{\infty} P_{i,k} = \frac{\bar{n}_1}{\mu_1} + \frac{\bar{n}_2}{\mu_2} . \quad (3.9)$$

Отримані співвідношення легко узагальнюються на випадок, коли потік запитів представляє собою суперпозицію m потоків з інтенсивностями надходження, які дорівнюють відповідно $\lambda_1, \lambda_2, \dots, \lambda_m$ та інтенсивностями обслуговування $\mu_1, \mu_2, \dots, \mu_m$. При цьому середнє число запитів j -го типу, що знаходяться в черзі дорівнює

$$\bar{n}_j = \frac{\rho_j}{1 - \rho_j}, j = 1, 2, \dots, m, \quad (3.10)$$

а середня тривалість очікування початку обслуговування для вхідного запиту, якій надійшов в систему в довільний момент часу, визначається співвідношенням

$$\bar{\tau} = \sum_{j=1}^m \frac{\bar{n}_j}{\mu_j} = \sum_{j=1}^m \bar{n}_j \cdot \bar{\theta}_j, \quad (3.11)$$

де $\bar{\theta}_j = \frac{1}{\mu_j}$ – середній час обслуговування запитів j -го типу.

3.3. Модель інтерфейсу вузла мережі з неоднорідним вхідним потоком пакетів з абсолютним пріоритетом.

Зрозуміло, що простота методу отримання наведених співвідношень є наслідком симетричності структури графа станів і переходів (рис. 3.2), що визначає динаміку станів в системі, коли потоки запитів рівноважні.

Разом з тим у багатьох практичних випадках необхідно, щоб запити будь-якого певного типу обслуговувалися в середньому швидше запитів інших типів. З метою забезпечення переважного права на обслуговування одних запитів перед іншими їм присвоюються свої пріоритети. Якщо пріоритет запиту враховується лише в момент вибору запита на обслуговування, то його називають відносним. При врахуванні цього пріоритету в момент звільнення сервісного каналу диспетчер вибирає з черги запит з найвищим пріоритетом, і він починає обслуговуватися. Відповідно до цього принципу робота вузла ТКС організовується таким чином, щоб запити, які мають різний пріоритет, шикувалися в свою чергу. В цьому випадку запити з черги з деяким конкретним пріоритетом будуть обслуговуватися тільки в тому випадку, якщо черги з більш високим пріоритетом не містять запитів. При цьому важливо зауважити, що запит, обслуговування якого почалося, продовжує обслуговування до закінчення, навіть якщо під час обслуговування надійшов запит з більш високим пріоритетом. Однак в деяких ситуаціях час очікування початку обслуговування для запитів деяких типів потрібно максимально знизити, що є недосяжним при використанні відносних пріоритетів. Практична можливість зменшення тривалості очікування таких запитів складається в організації роботи системи таким чином, щоб такий високопріоритетний запит починав обслуговуватися негайно в момент його надходження, перериваючи розпочате обслуговування запиту, якщо він має більш низький пріоритет. Відповідну дисципліну обслуговування називають обслуговуванням з абсолютним пріоритетом.

Проведемо аналіз ефективності функціонування вузла ТКС з урахуванням відмінностей в пріоритетах запитів [7].

Нехай на вхід вузла ТКС надходить суперпозиція двох потоків запитів, причому запити першого потоку мають абсолютний пріоритет перед запитами другого потоку. Граф станів і переходів для системи такого виду наведено на малюнку 3.6.

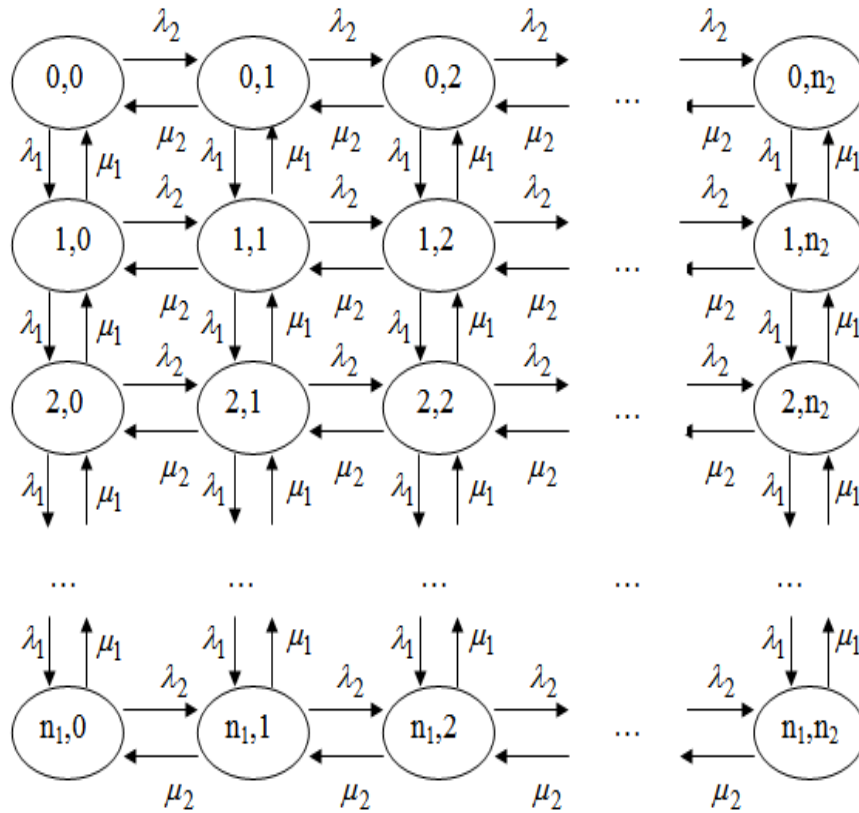


Рис. 3.6. Граф станів і переходів для системи з пріоритетом.

Відображенням на графі наявності абсолютного пріоритету запитів першого типу перед запитами другого типу є відсутність дуг, що пов'язують вершини (i, j) і $(i, j-1)$ при $i \neq 0$, відповідних обслуговуванню запитів другого типу при наявності в системі більш пріоритетних запитів першого типу.

Проведемо аналіз такої системи. Здійснимо укрупнення станів цього графа, об'єднуючи, з урахуванням специфіки цього графа, в групи стану наступним чином:

$$E_{\Gamma 0} = \{E_{00}, E_{10}, E_{20}, \dots, E_{n_1 0}\}, E_{\Gamma 1} = \{E_{01}, E_{11}, E_{21}, \dots, E_{n_1 1}\}, \dots, E_{\Gamma n_2} = \{E_{0n_2}, E_{1n_2}, E_{2n_2}, \dots, E_{n_1 n_2}\}$$

В результаті укрупнення станів еквівалентний граф станів і переходів буде мати вигляд, представлений на малюнку 3.7.

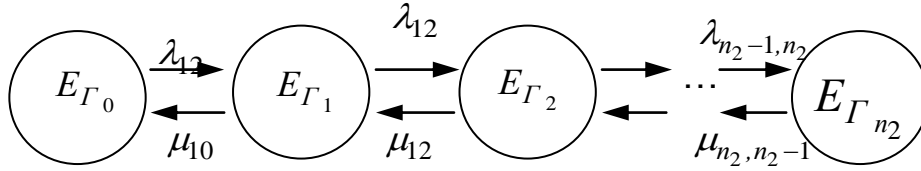


Рис. 3.7. Граф станів і переходів для системи з укрупненим станом. Інтенсивності переходів в цій схемі розраховуються за формулами

$$\begin{aligned}\lambda_{k,k+1} &= \sum_{i=0}^{n_2} \lambda_2 \hat{P}_{ik} = \lambda_2, k=0,1,\dots,n_2-1 \\ \mu_{k,k+1} &= \hat{P}_{0,k+1} \mu_2, k=0,1,\dots,n_2-1\end{aligned}\quad (3.12)$$

тут \hat{P}_{ik} – умовна ймовірність перебування в i -ому стані k -го шару, за умови, що система знаходиться в одному з станів цього шару.

Введемо ймовірність $P_{\Gamma k}$, $k=0,1,\dots,n_2$, перебування системи в k -ому шарі. За аналогією з попереднім завданням складемо систему рівнянь щодо невідомих ймовірностей $P_{\Gamma k}$, $k=0,1,\dots,n_2$. З урахуванням (3.12) маємо

$$\begin{cases} \mu_2 \hat{P}_{01} P_{\Gamma 1} - \lambda_2 P_{\Gamma 0} = 0, \\ \mu_2 \hat{P}_{02} P_{\Gamma 2} + \lambda_2 P_{\Gamma 0} - (\mu_2 \hat{P}_{01} + \lambda_2) P_{\Gamma 1} = 0, \\ \dots\dots\dots \\ \lambda_2 P_{\Gamma,k-1} + \mu_2 \hat{P}_{0,k+1} P_{\Gamma,k+1} - (\lambda_2 + \mu_2 \hat{P}_{0k}) P_{\Gamma k} = 0, \\ \dots\dots\dots \\ \lambda_2 P_{\Gamma,n_2-1} - \mu_2 \hat{P}_{0,n_2} \hat{P}_{\Gamma,n_2} = 0. \end{cases}\quad (3.13)$$

Введемо $u_k = \lambda_2 P_{\Gamma,k-1} - \mu_2 \hat{P}_{0,k} P_{\Gamma k}$. Тоді (3.13) перетворюється до вигляду:

$$\begin{aligned}u_1 &= 0, \\ u_1 - u_2 &= 0, \\ \dots\dots\dots \\ u_k - u_{k+1} &= 0, \\ u_{n_2} &= 0,\end{aligned}$$

Звідси $u_1 = u_2 = \dots = u_k = \dots u_{n_2} = 0$, тобто $\lambda_2 P_{\Gamma, k-1} - \mu_2 \hat{P}_{0, k} P_{\Gamma k} = 0$.

Звідси випливає рекурентне співвідношення

$$P_{\Gamma k} = \frac{\lambda_2}{\mu_2 \hat{P}_{0, k}} P_{\Gamma, k-1}, \quad k = 1, 2, \dots, n_2 - 1$$

При цьому

$$P_{\Gamma 1} = \frac{\lambda_2}{\mu_2 \hat{P}_{0, 1}} P_{\Gamma, 0} = \frac{\rho_2}{\hat{P}_{0, 1}} P_{\Gamma, 0},$$

$$P_{\Gamma 2} = \frac{\lambda_2}{\mu_2 \hat{P}_{0, 2}} P_{\Gamma, 1} = \frac{\rho_2^2}{\hat{P}_{0, 1} \hat{P}_{0, 2}} P_{\Gamma, 0},$$

.....

$$P_{\Gamma k} = \frac{\lambda_2}{\mu_2 \hat{P}_{0, k}} P_{\Gamma, k-1} = \frac{\rho_2^k}{\hat{P}_{0, 1} \hat{P}_{0, 2} \dots \hat{P}_{0, k}} P_{\Gamma, 0},$$

$$P_{\Gamma n_2} = \frac{\rho_2^{n_2}}{\hat{P}_{0, 1} \hat{P}_{0, 2} \dots \hat{P}_{0, n_2}}.$$

Таким чином, всі невідомі ймовірності $P_{\Gamma 1}, P_{\Gamma 2}, \dots, P_{\Gamma n_2}$ виражені через ймовірність $P_{\Gamma 0}$, яку знайдемо з умови нормування

$$\sum_{k=0}^{n_2} P_{\Gamma k} = \left(\sum_{k=0}^{n_2} \frac{\rho_2^k}{\prod_{l=1}^{n_2} \hat{P}_{0l}} \right) P_{\Gamma 0} = 1.$$

Звідси

$$P_{\Gamma 0} = \frac{1}{\sum_{k=0}^{n_2} \frac{\rho_2^k}{\prod_{l=1}^{n_2} \hat{P}_{0l}}} \quad (3.14)$$

та

$$P_{\Gamma k} = \frac{\frac{\rho_2^k}{\prod_{l=1}^k \hat{P}_{0l}}}{\sum_{k=0}^{n_2} \frac{\rho_2^k}{\prod_{l=1}^k \hat{P}_{0l}}}, k = 0, 1, \dots, n_2. \quad (3.15)$$

Для пошуку умовних ймовірностей $\hat{P}_{00}, \hat{P}_{01}, \dots, \hat{P}_{0k}, \hat{P}_{0n_2}$ використовуємо ту обставину, що система переходів і значення інтенсивностей переходів в групах станів $E_{\Gamma 0}, E_{\Gamma 1}, \dots, E_{\Gamma n_1}$ в графі на рис.3.7 нічим не відрізняються від відповідних значень в графі, представленою на рис.3.5. Тому, з урахуванням (3.3), запишемо

$$\hat{P}_{0k} = \frac{1 - \rho_1}{1 - \rho_1^{n_1+1}}, k = 0, 1, \dots, n_1,$$

тоді

$$\prod_{l=0}^k \hat{P}_{0l} = \frac{(1 - \rho_1)^k}{(1 - \rho_1^{n_1+1})^k} \quad (3.16)$$

Підставляючи (3.16) в (3.15), отримуємо

$$P_{\Gamma k} = \frac{\left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^k}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^l}, k = 0, 1, \dots, n. \quad (3.17)$$

Тепер неважко отримати безумовні ймовірності P_{ik} , $i=0,1,...,n_1$, $k=0,1,...,n_2$ станів системи. Оскільки для кожного набору станів $\{E_{0k}, E_{1k}, ..., E_{n_1,k}\}$, $k=0,1,...,n_2$, відповідно до (3.3) має місце

$$\hat{P}_{ik} = \frac{\rho_1^i}{\sum_{l=0}^{n_1} \rho_1^l} = \frac{\rho_1^i (1 - \rho_1)}{1 - \rho_1^{n_1+1}},$$

де \hat{P}_{ik} – умовна ймовірність перебування системи в i -ому стані безлічі $\{E_{0k}, E_{1k}, ..., E_{n_1,k}\}$, то безумовні ймовірності P_{ik} визначаються співвідношенням

$$P_{ik} = \hat{P}_{ik} P_{\Gamma k} = \frac{\frac{\rho_1^i (1 - \rho_1)}{1 - \rho_1^{n_1+1}} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^k}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^l} = \frac{\rho_1^i \rho_2^k (1 - \rho_1^{n_1+1})^{k-1}}{1 - \rho_1} \cdot \frac{1}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^l}, \quad (3.18)$$

$i = 0, 1, ..., n_1, \quad k = 0, 1, ..., n_2$

Якщо довжини черг для запитів обох типів необмежені, то співвідношення (3.18) трохи спрощуються. При цьому необхідно мати на увазі,

що $\lim_{n \rightarrow \infty} \rho_1^{n+1} = 0$, а ряд $1, \frac{\rho_2}{1 - \rho_1}, \left(\frac{\rho_2}{1 - \rho_1} \right)^2, ...,$ як легко побачити, сходиться тільки в тому випадку, якщо $\rho_1 + \rho_2 < 1$. В цьому випадку наведена послідовність є нескінченно спадною геометричною прогресією і

$$\sum_{l=0}^{\infty} \left(\frac{\rho_2}{1 - \rho_1} \right)^l = \frac{1}{1 - \frac{\rho_2}{1 - \rho_1}} = \frac{1 - \rho_1}{1 - \rho_2 - \rho_1}.$$

Тоді

$$P_{ik} = \rho_1^i \left(\frac{\rho_2}{1-\rho_1} \right)^k (1-\rho_1-\rho_2), \quad i=0,1,2,\dots, \quad k=0,1,2,\dots \quad (3.19)$$

Використовуючи отримане співвідношення, розрахуємо середні значення кількості запитів першого типу \bar{n}_1 і другого типу \bar{n}_2 , які очікують у черзі:

$$\begin{aligned} \bar{n}_1 &= \sum_{i=0}^{\infty} i \sum_{k=0}^{\infty} P_{ik} = \sum_{i=0}^{\infty} i \sum_{k=0}^{\infty} \rho_1^i \left(\frac{\rho_2}{1-\rho_1} \right)^k (1-\rho_1-\rho_2) = \\ &= (1-\rho_1-\rho_2) \sum_{i=0}^{\infty} i \rho_1^i \sum_{k=0}^{\infty} \left(\frac{\rho_2}{1-\rho_1} \right)^k = (1-\rho_1) \sum_{i=1}^{\infty} \rho_1^i = \frac{\rho_1}{1-\rho_1}. \end{aligned} \quad (3.20)$$

Отриманий результат, природно, збігається з (3.10), оскільки запити другого типу в силу більш низького пріоритету ніяк не впливають на процес обслуговування запитів першого типу. Далі

$$\begin{aligned} \bar{n}_2 &= \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} k P_{ik} = \sum_{i=0}^{\infty} \sum_{k=1}^{\infty} k \rho_1^i \left(\frac{\rho_2}{1-\rho_1} \right)^k (1-\rho_1-\rho_2) = \\ &= (1-\rho_1-\rho_2) \sum_{i=0}^{\infty} i \rho_1^i \sum_{k=1}^{\infty} \left(\frac{\rho_2}{1-\rho_1} \right)^k = \frac{1-\rho_1-\rho_2}{1-\rho_1} \sum_{k=1}^{\infty} \frac{k \rho_2}{1-\rho_1} \left(\frac{\rho_2}{1-\rho_1} \right)^{k-1} = \\ &= \frac{(1-\rho_1-\rho_2) \rho_2}{1-\rho_1} \sum_{k=1}^{\infty} \frac{d}{d \rho_2} \left(\frac{\rho_2}{1-\rho_1} \right)^k = \frac{(1-\rho_1-\rho_2) \rho_2}{1-\rho_1} \frac{d}{d \rho_2} \sum_{k=1}^{\infty} \frac{d}{d \rho_2} \left(\frac{\rho_2}{1-\rho_1} \right)^k = \\ &= \frac{(1-\rho_1-\rho_2) \rho_2}{1-\rho_1} \frac{d}{d \rho_2} \left(\frac{1-\rho_1}{1-\rho_2-\rho_1} \right) = \frac{(1-\rho_1-\rho_2) \rho_2}{1-\rho_1} \frac{1-\rho_1}{(1-\rho_2-\rho_1)^2} = \frac{\rho_2}{1-\rho_2-\rho_1} \end{aligned} \quad (3.21)$$

Як слід було очікувати, середня довжина черги для запитів другого типу залежить від інтенсивності потоку запитів першого типу і продуктивності їх обробки.

При цьому середній час очікування початку обслуговування для запитів першого типу дорівнює

$$\bar{\tau}_1 = \sum_{i=1}^{\infty} \frac{i}{\mu_1} \sum_{k=0}^{\infty} P_{ik} = \frac{\bar{n}_1}{\mu_1},$$

а для запитів другого типу визначається співвідношенням

$$\bar{\tau}_2 = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \left(\frac{i}{\mu_1} + \frac{k}{\mu_2} \right) P_{ik} = \frac{\bar{n}_1}{\mu_1} + \frac{\bar{n}_2}{\mu_2}.$$

Запишемо, нарешті, співвідношення для сумарної довжини черги

$$\begin{aligned} n = n_1 + n_2 &= \frac{\rho_1}{1 - \rho_1} + \frac{\rho_2}{1 - \rho_1 - \rho_2} = \frac{\rho_1 - \rho_1^2 - \rho_1\rho_2 + \rho_2 - \rho_1\rho_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} = \\ &= \frac{(\rho_1 + \rho_2) - (\rho_1 + \rho_2)^2 + \rho_2^2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} = \frac{(\rho_1 + \rho_2)(1 - \rho_1 - \rho_2) + \rho_2^2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} = \\ &= \frac{\rho_1 + \rho_2}{1 - \rho_1} + \frac{\rho_2^2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} \end{aligned} \quad (3.22)$$

3.4. Модель багатоканального вузла мережі з неоднорідним вхідним потоком пакетів

Аналіз багатоканальних вузлів ТКС з неоднорідним вхідним потоком запитів [8] проводиться аналогічно попередньому. Як і раніше, з метою простоти пояснення і спрощення запису рішення завдання аналізу багатоканальної мережі проведено для окремого випадку, коли на вхід мережі надходить суперпозиція двох потоків запитів, причому запити першого потоку мають абсолютний пріоритет перед запитами другого потоку.

Введемо безліч можливих станів системи $\{E_{ij}\}, i = 1, 2, \dots, n_1, j = 1, 2, \dots, n_2$.

Тут E_{ij} – стан, відповідний ситуації, коли в системі знаходяться i запитів першого типу і j запитів другого. Зауважимо, що для багатоканальної системи трактування індексів i і j в позначенні стану різне в залежності від чисельних значень цих індексів. Нехай мережа містить два канали обробки запитів. Перерахуємо можливі варіанти трактування:

$(i=0, j \leq 2)$ – немає запитів першого типу, j запитів другого типу обслуговуються;

$(i=0, j > 2)$ – немає запитів першого типу, два запита другого типу обслуговуються, $j-2$ запитів другого типу в – черзі;

$(i=1, j \leq 1)$ – один запит першого типу обслуговується, j запитів другого типу обслуговуються;

$(i=1, j > 1)$ – один запит першого типу обслуговується, один запит другого типу обслуговується, $j-1$ запитів другого типу – в черзі;

$(i=2, j)$ – два запита першого типу обслуговуються, j запитів другого типу – в черзі;

$(i > 2, j)$ – два запита першого типу обслуговуються, $i-2$ запитів першого типу – в черзі, j запитів другого типу – в черзі.

Для ℓ –канальної системи з пріоритетом:

$(i \leq \ell, j)$ – i запитів першого типу обслуговуються, $(\ell-i)$ запитів другого типу обслуговуються, $j-(\ell-i)=j+i-\ell$ запитів другого типу – в черзі;

$(i > \ell, j)$ – ℓ запитів першого типу обслуговуються, $(i-\ell)$ запитів першого типу – в черзі, j запитів другого типу – в черзі.

Нехай на вхід надходить два пуассонівських потоки запитів з інтенсивностями λ_1 і λ_2 та інтенсивностями обслуговування μ_1 і μ_2 . Граф станів і переходів такої системи представлений на рис. 3.8.



Відповідно до технології фазового укрупнення введемо сукупність групових станів (шарів).

Граф станів і переходів для такої укрупненої системи наведено на рис.

68

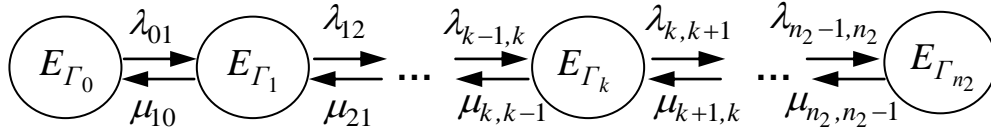


Рис. 3.9. Граф станів та переходів після укрупнення станів.

Нехай $\lambda_{k-1,k}$ – інтенсивність переходів з укрупненого стану $E_{\Gamma_{k-1}}$ в укрупнений стан E_{Γ_k} , а $\mu_{k,k-1}$ – інтенсивність переходу з E_{Γ_k} в $E_{\Gamma_{k-1}}$. Введемо розподіл ймовірностей укрупнених станів $\{P_{\Gamma_0}, P_{\Gamma_1}, \dots, P_{\Gamma_k}, \dots, P_{\Gamma_{n_2}}\}$.

Тоді система рівнянь Колмогорова щодо цих ймовірностей має вигляд:

$$\begin{aligned} \mu_{10}P_{\Gamma_1} - \lambda_{01}P_{\Gamma_0} &= 0, \\ \lambda_{01}P_{\Gamma_0} + \mu_{10}P_{\Gamma_2} - (\lambda_{12} + \mu_{10})P_{\Gamma_1} &= 0, \\ \lambda_{k-1,k}P_{\Gamma_{k-1}} + \mu_{k+1,k}P_{\Gamma_{k+1}} - (\lambda_{k,k+1} + \mu_{k,k-1})P_{\Gamma_k} &= 0, \\ \lambda_{n_2-1,n_2}P_{\Gamma_{n_2-1}} - \mu_{n_2,n_2-1}P_{\Gamma_{n_2}} &= 0 \end{aligned} \quad (3.23)$$

Введемо нову змінну $u_k = \lambda_{k-1,k}P_{\Gamma_{k-1}} - \mu_{k,k-1}P_{\Gamma_k}$.

Тоді система (3.23) прийме вигляд:

$$\begin{aligned} u_1 &= 0, \\ u_k - u_{k+1} &= 0, k = 1, 2, \dots, n_2 - 1, \\ u_{n_2} &= 0 \end{aligned}$$

звідси $u_1 = u_2 = \dots = u_k = u_{n_2} = 0$.

При цьому

$$\begin{aligned} \lambda_{k-1,k}P_{\Gamma_{k-1}} - \mu_{k,k-1}P_{\Gamma_k} &= 0 \\ P_{\Gamma_k} &= \frac{\lambda_{k-1,k}}{\mu_{k,k-1}}P_{\Gamma_{k-1}}, k = 1, 2, \dots, n_2 \end{aligned} \quad (3.24)$$

З рекурентного співвідношення (3.24) виходить:

$$\begin{aligned}
 P_{\Gamma_1} &= \frac{\lambda_{01}}{\mu_{10}} P_{\Gamma_0}, \\
 P_{\Gamma_2} &= \frac{\lambda_{12}}{\mu_{21}} P_{\Gamma_1} = \frac{\lambda_{01}\lambda_{12}}{\mu_{10}\mu_{21}} P_{\Gamma_0}, \\
 &\dots\dots\dots \\
 P_{\Gamma_k} &= \frac{\lambda_{k-1,k}}{\mu_{k,k-1}} P_{\Gamma_{k-1}} = \prod_{l=1}^k \frac{\lambda_{l-1,l}}{\mu_{l,l-1}} P_{\Gamma_0}, \\
 &\dots\dots\dots \\
 P_{\Gamma_{n_2}} &= P_{\Gamma_0} \prod_{l=1}^{n_2} \frac{\lambda_{l-1,l}}{\mu_{l,l-1}}
 \end{aligned} \tag{3.25}$$

Оскільки $\sum_{k=0}^{n_2} P_{\Gamma_k} = P_{\Gamma_0} \left(1 + \sum_{k=1}^{n_2} \prod_{l=1}^k \frac{\lambda_{l-1,l}}{\mu_{l,l-1}} \right) = I$, тоді $P_{\Gamma_0} = \frac{I}{1 + \sum_{k=1}^{n_2} \prod_{l=1}^k \frac{\lambda_{l-1,l}}{\mu_{l,l-1}}}$.

Тоді

$$P_{\Gamma_k} = \frac{\prod_{l=1}^k \frac{\lambda_{l-1,l}}{\mu_{l,l-1}}}{1 + \sum_{k=1}^{n_2} \prod_{l=1}^k \frac{\lambda_{l-1,l}}{\mu_{l,l-1}}}, \quad k = 1, 2, \dots, n_2 \tag{3.26}$$

Введемо

$$P_l = \frac{\lambda_{l-1,l}}{\mu_{l,l-1}}, \quad l = 1, 2, \dots, n_2. \tag{3.27}$$

При цьому співвідношення (3.26) прийме вигляд

$$P_{\Gamma_0} = \frac{I}{1 + \sum_{k=1}^{n_2} \prod_{l=1}^k P_l}, \quad P_{\Gamma_k} = \frac{\prod_{l=1}^k P_l}{1 + \sum_{k=1}^{n_2} \prod_{l=1}^k P_l}, \quad k = 1, 2, \dots, n_2. \tag{3.28}$$

Знайдемо $P_{l, l=1,2,\dots,n_2}$. Значення $\lambda_{l-1,l}$, $\mu_{l,l-1}$, $l = 1, 2, \dots, n_2$, залежать від числа сервісних каналів. Якщо, в найпростішому випадку, $s = 2$, тоді

$$\begin{aligned}\lambda_{l-1,l} &= \sum_{j=0}^{n_l} \lambda_2 \hat{P}_{i,l-1} = \lambda_2 \sum_{i=0}^{n_l} \hat{P}_{i,l-1} = \lambda_2, \quad l = 1, 2, \dots, n_2, \\ \mu_{10} &= (\hat{P}_{01} + \hat{P}_{11}) \mu, \\ \mu_{l,l-1} &= 2\mu_2 \hat{P}_{0l} + \mu_2 \hat{P}_{1l} = (2\hat{P}_{0l} + \hat{P}_{1l}) \mu_2, \quad l = 2, 3, \dots, n_2.\end{aligned}\tag{3.29}$$

У цих співвідношеннях \hat{P}_{ik} є умовна ймовірність перебування системи в i -ому стані k -го шару, за умови, що система знаходиться в цьому шарі. Оскільки дисципліна обслуговування пріоритетних запитів першого типу не залежить від числа запитів другого типу в системі, то

$$\begin{aligned}\hat{P}_{00} &= \hat{P}_{01} = \hat{P}_{02} = \dots = \hat{P}_{0,n_2}, \\ \hat{P}_{i0} &= \hat{P}_{i1} = \hat{P}_{i2} = \dots = \hat{P}_{i,n_2} \quad i = 1, 2, \dots, n_2\end{aligned}\tag{3.30}$$

Це позначає, що для відшукування цих ймовірностей можливо використовувати будь-яку з груп станів $E_{\Gamma_0}, E_{\Gamma_1}, \dots, E_{\Gamma_{n_2}}$, наприклад, групу E_{Γ_1} .

Тому

$$\mu_{10} = (\hat{P}_{01} + \hat{P}_{11}) \mu_2, \quad \mu_{l,l-1} = (2\hat{P}_{0l} + \hat{P}_{1l}) \mu_2\tag{3.31}$$

Ймовірності \hat{P}_{0l} , \hat{P}_{1l} розрахуємо, використовуючи систему рівнянь для ймовірностей станів E_{Γ_i} .

$$\begin{aligned}\mu_1 \hat{P}_{11} - \lambda_1 \hat{P}_{01} &= 0, \\ \mu_1 \hat{P}_{i+1,l} - \lambda_1 \hat{P}_{i-1,l} - (\lambda_l + \mu_l) \hat{P}_{i,l} &= 0, \quad i = 1, 2, \dots, n_l\end{aligned}\tag{3.32}$$

Введемо

$$\begin{aligned} V_1 &= \lambda_1 \hat{P}_{01} - \mu_1 \hat{P}_{11}, \\ V_i &= \lambda_1 \hat{P}_{i-1,1} - 2\mu_1 \hat{P}_{i,1}, \quad i = 1, 2, \dots, n_1 \end{aligned}$$

З врахуванням цих позначень система (3.32) прийме вигляд,

$$\begin{aligned} V_1 &= 0, \\ V_i - V_{i+1} &= 0, \quad k = 1, 2, \dots, n_1, \\ n_1 &= 0. \end{aligned}$$

Звідси $V_1 = V_2 = \dots = V_i = V_{n_1} = 0$.

Тоді

$$\begin{aligned} \hat{P}_{11} &= \frac{\lambda_1}{\mu_1} \hat{P}_{01}, \\ \hat{P}_{21} &= \frac{\lambda_1}{2\mu_1} \hat{P}_{11} = \frac{\lambda_1^2}{2\mu_1^2} \hat{P}_{01}, \\ \hat{P}_{i,1} &= \frac{\lambda_1^i}{2^{i-1} \mu_1^i} \hat{P}_{01}, \quad i = 1, 2, \dots, n_1. \end{aligned}$$

Значення \hat{P}_{01} знайдемо з умови нормування. При цьому

$$\begin{aligned} \hat{P}_{01} \left(1 + \sum_{i=1}^{n_1} \frac{\lambda_1^i}{2^{i-1} \mu_1^i} \right) &= 1, \quad \hat{P}_{01} = \frac{1}{1 + \sum_{i=1}^{n_1} \frac{\lambda_1^i}{2^{i-1} \mu_1^i}}, \\ \hat{P}_{01} &= \frac{\frac{\lambda_1}{\mu_1}}{1 + \sum_{i=1}^{n_1} \frac{\lambda_1^i}{2^{i-1} \mu_1^i}}, \quad \hat{P}_{01} = \frac{\frac{\lambda_1^i}{2^{i-1} \mu_1^i}}{1 + \sum_{i=1}^{n_1} \frac{\lambda_1^i}{2^{i-1} \mu_1^i}}, \quad i = 1, 2, \dots, n_1. \end{aligned}$$

Введемо $\alpha_l = \frac{\lambda_l}{\mu_l}$. Тоді

$$\hat{P}_{0l} = \frac{1}{1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}}}, \quad \hat{P}_{0l} = \frac{\frac{\alpha_l^i}{2^{i-1}}}{1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}}}, \quad i = 1, 2, \dots, n_l \quad (3.33)$$

Відтепер у відповідності з (3.31),

$$\mu_{10} = \frac{(1 + \alpha_l)\mu_2}{1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}}}, \quad \mu_{l,l-1} = \frac{(2 + \alpha_l)\mu_2}{1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}}}. \quad (3.34)$$

Підставляючи (3.30), (3.34) в (3.27), отримуємо

$$\rho = \frac{\lambda_{0l}}{\mu_{10}} = \frac{\lambda_2 \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)}{\mu_2 (1 + \alpha_l)},$$

$$\rho = \frac{\lambda_{l-1,l}}{\mu_{l,l-1}} = \frac{\lambda_2 \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)}{\mu_2 (2 + \alpha_l)}, \quad l = 2, 3, \dots, n_2.$$

Введемо $\alpha_2 = \frac{\lambda_2}{\mu_2}$. Тоді

$$\rho = \frac{\alpha_2 \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)}{1 + \alpha_l}, \quad (3.35)$$

$$\rho_l = \frac{\alpha_2 \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)}{1 + \alpha_l}, \quad l = 2, 3, \dots, n_2. \quad (3.36)$$

Тепер, підставляючи (3.35), (3.36) в (3.28), отримаємо співвідношення для ймовірнісних групових станів.

$$P_{\Gamma_k} = \frac{\frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)^k}{(1 + \alpha_l)(2 + \alpha_l)^{k-1}}}{1 + \sum_{k=1}^{n_2} \frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)^k}{1 + \alpha_l}}, \quad k = 1, 2, \dots, n_2, \quad (3.37)$$

$$P_{\Gamma_0} = \frac{1}{\frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)^k}{(1 + \alpha_l)(2 + \alpha_l)^{k-1}}}. \quad (3.38)$$

Отримані співвідношення дозволяють розрахувати ймовірності станів системи. Так, якщо $P_{ik} = P_{ik} P_{\Gamma_k}$, то, з урахуванням (3.30), (3.33), (3.37), (3.38), запишемо

$$P_{00} = \frac{1}{\left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right) \left(\frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_l} \frac{\alpha_l^i}{2^{i-1}} \right)^k}{(1 + \alpha_l)(2 + \alpha_l)^{k-1}} \right)}, \quad (3.39)$$

$$P_{00} = \frac{\frac{\alpha_1^i}{2^{i-1}}}{\left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right) \left(\frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 + \alpha_1)(2 + \alpha_1)^{k-1}} \right)}, \quad i = 1, 2, \dots, n_1. \quad (3.40)$$

$$P_{0k} = \frac{\frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 + \alpha_1)(2 + \alpha_1)^{k-1}}}{\left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right) \left(\frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 + \alpha_1)(2 + \alpha_1)^{k-1}} \right)}, \quad k = 1, 2, \dots, n_2. \quad (3.41)$$

$$P_{ik} = \frac{\frac{\alpha_1^i}{2^{i-1}} \cdot \frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 - \alpha_1)(2 + \alpha_1)^{k-1}}}{\left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right) \left(1 + \sum_{k=1}^{n_2} \frac{\alpha_2^k \left(1 + \sum_{i=1}^{n_1} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 - \alpha_1)(2 + \alpha_1)^{k-1}}\right)}, \quad i = 1, 2, \dots, n_1; \quad k = 1, 2, \dots, n_2. \quad (3.42)$$

Якщо довжина черг не обмежена, то співвідношення (3.39) – (3.42) спрощуються. Як буде зрозуміло з подальшого, довжина жодної з черг не буде зростати до нескінченності, якщо $\alpha_1 + \alpha_2 < 2$.

Тоді, так як в цьому випадку $\alpha_1 < 2$, то

$$1 + \sum_{i=1}^{\infty} \frac{\alpha_1^i}{2^{i-1}} = 1 + 2 \sum_{i=1}^{\infty} \left(\frac{\alpha_1}{2}\right)^i = 1 + 2 \frac{\frac{\alpha_1}{2}}{1 - \frac{\alpha_1}{2}} = \frac{2\alpha_1}{2 - \alpha_1} + 1 = \frac{2 + \alpha_1}{2 - \alpha_1}. \quad (3.43)$$

При цьому, після нескладних перетворень, отримаємо

$$\frac{\alpha_2^k \left(1 + \sum_{i=1}^{\infty} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 + \alpha_1)(2 + \alpha_1)^{k-1}} = \frac{2 + \alpha_1}{1 + \alpha_1} \left(\frac{\alpha_2}{2 - \alpha_1}\right)^2. \quad (3.44)$$

$$1 + \sum_{k=1}^{\infty} \frac{\alpha_2^k \left(1 + \sum_{i=1}^{\infty} \frac{\alpha_1^i}{2^{i-1}}\right)^k}{(1 + \alpha_1)(2 + \alpha_1)^{k-1}} = 1 + \frac{2 + \alpha_1}{1 + \alpha_1} \sum_{i=1}^{\infty} \left(\frac{\alpha_2}{2 - \alpha_1}\right)^k = \frac{2 + \alpha_1 - \alpha_1^2 + \alpha_2}{(1 + \alpha_1)(2 - \alpha_1 - \alpha_2)}. \quad (3.45)$$

Тепер, з урахуванням (3.43), (3.44), (3.45), запишемо співвідношення (3.39) – (3.42) наступним чином:

$$\begin{aligned} P_{00} &= \frac{1}{\frac{2 + \alpha_1}{1 + \alpha_1} \cdot \frac{2 + \alpha_1 - \alpha_1^2 + \alpha_2}{(1 + \alpha_1)(2 - \alpha_1 - \alpha_2)}} = \frac{2 + \alpha_1 - \alpha_1^2 + \alpha_2}{(1 + \alpha_1)(2 - \alpha_1 - \alpha_2)}, \\ P_{i0} &= \frac{\alpha_1^i (2 - \alpha_1)(1 + \alpha_1)(2 - \alpha_1 - \alpha_2)}{2^{i-1} (2 + \alpha_1)(2 + \alpha_1 - \alpha_1^2 + \alpha_2)}, \\ P_{0k} &= \frac{\alpha_2^k (2 - \alpha_1 - \alpha_2)}{(2 + \alpha_1)^{k-1} (2 + \alpha_1 - \alpha_1^2 + \alpha_2)}, \\ P_{ik} &= \frac{\alpha_1^i \alpha_2^k (2 - \alpha_1 - \alpha_2)}{2^{i-1} (2 + \alpha_1)^{k-1} (2 + \alpha_1 - \alpha_1^2 + \alpha_2)}. \end{aligned} \quad (3.46)$$

Отримані співвідношення (3.46) забезпечують можливість розрахунку розподілу ймовірностей станів двоканального вузла телекомунікаційної мережі, на вхід якої надходять два потоки запитів з пріоритетом першого перед другим. Цілком аналогічно проводиться аналіз багатоканальної мережі.

3.5. Модель вузла маршрутизації з неоднорідним вхідним потоком пакетів з відносним пріоритетом

Розглянемо тепер більш складну дисципліну обслуговування, що виникає при наявності відносного пріоритету одних запитів перед іншими. Як вже було зазначено, в цій ситуації запит, що надходить в момент, коли система обслуговує інший запит з більш низьким пріоритетом, не перериває його обробку, а стає в свою чергу запитів і очікує початку обслуговування. У зв'язку з цим, якщо в системі (для випадку двох потоків) знаходяться i запитів з більш високим пріоритетом і j запитів з низьким пріоритетом, необхідно розрізняти стан (i, j) , в якому обслуговується запит першого типу з високим пріоритетом, від стану (i, j) , коли обслуговується запит другого типу. При цьому в першому випадку після закінчення обслуговування запиту з інтенсивністю μ_1 система переходить в стан $(i-1, j)$, а другому – з інтенсивністю μ_2 переходить в стан $(i, j-1)$. Для розрізнення цих двох станів будемо відзначати штрихом індекс стану i або j відповідно до того, запит якого саме типу обслуговується. При цьому стан (i', j) описує ситуацію, коли обслуговується запит першого типу, а стан (i, j') – другого типу.

З урахуванням сказаного, граф станів і переходів для вузла телекомунікаційної мережі, на вхід якого надходить суперпозиція двох потоків запитів, з яких перший потік запитів має відносний пріоритет перед другим зображений на рис. 3.10.

Для аналізу системи введемо групові стани наступним чином:

$$E_{\Gamma_0} = \left\{ E_{00}, E_{01}, \dots, E_{0n_2}; E_{1'0}, E_{1'1}, \dots, E_{1'n_2}; \dots E_{n_1'0}, \dots, E_{n_1'n_2} \right\},$$

$$E_{\Gamma_1} = \left\{ E_{11'}, E_{12'}, \dots, E_{1n_2'}; E_{21'}, \dots, E_{2n_2'}; \dots E_{n_11'}, \dots, E_{n_1n_2'} \right\}.$$

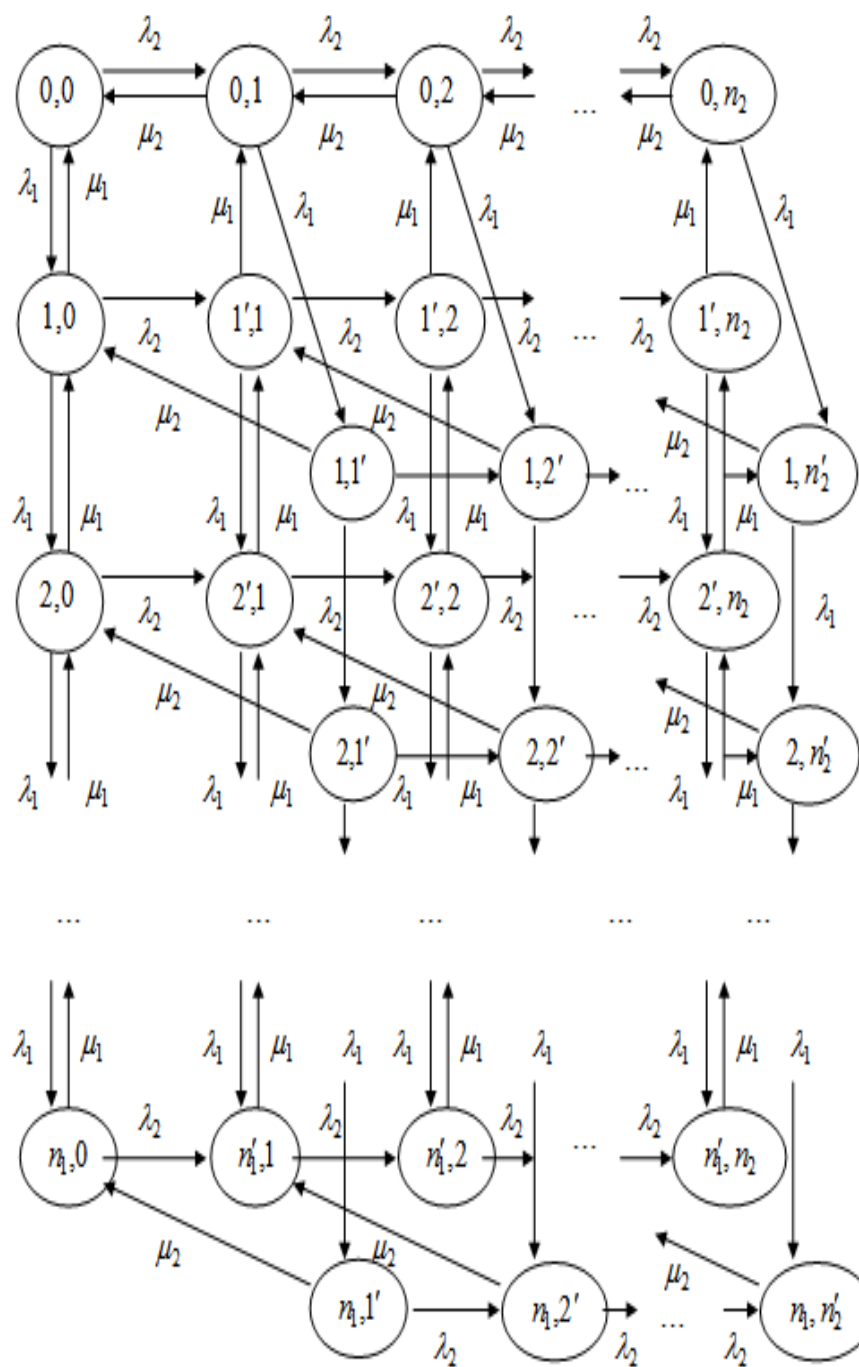


Рис. 3.10. Граф станів і переходів для системи з відносним пріоритетом

При цьому граф укрупнених станів і переходів має вигляд, наведений на рис. 3.11.

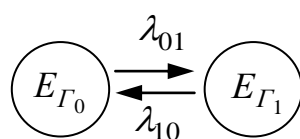


Рис. 3.11. Граф групових станів та переходів.

З рис. 3.11 зрозуміло, що

$$\begin{aligned}\lambda_{01} &= \lambda_1 \sum_{i=1}^{n_2} P'_{0i}, \\ \lambda_{01} &= \mu_2 \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} P'_{i_1 i_2}.\end{aligned}\tag{3.47}$$

Тут

P'_{0i_2} – умовна ймовірність стану E_{0i_2} ,

$P'_{i_1 i_2}$ – умовна ймовірність стану $E_{i_1 i_2}$.

Тоді

$$P_{\Gamma_1} \lambda_{10} - P_{\Gamma_0} \lambda_{01} = 0,$$

звідси

$$P_{\Gamma_1} = P_{\Gamma_0} \frac{\lambda_{01}}{\lambda_{10}}.$$

Оскільки

$$P_{\Gamma_1} + P_{\Gamma_0} = 0,$$

тоді

$$P_{\Gamma_0} \left(\frac{\lambda_{01}}{\lambda_{10}} + 1 \right) = P_{\Gamma_0} \frac{\lambda_{01} + \lambda_{10}}{\lambda_{10}} = 1,$$

$$P_{r_0} = \frac{\lambda_{10}}{\lambda_{01} + \lambda_{10}},$$

$$P_{\Gamma_1} = \frac{\lambda_{01}}{\lambda_{01} + \lambda_{10}}.$$

При цьому безумовні ймовірності станів системи визначаються співвідношеннями

$$P_{i_{i_2}} = P'_{i_{i_2}} \frac{\lambda_{01}}{\lambda_{10} + \lambda_{01}}, \quad (3.48)$$

$$P_{i_1 i_2'} = P'_{i_1 i_2'} \frac{\lambda_{01}}{\lambda_{01} + \lambda_{10}}$$

де P'_{ij_2} і $P'_{ij'_2}$ – умовні ймовірності відповідних станів в шарах Γ_0 і Γ_1 .

Умовний розподіл ймовірностей для групового стану Γ_0 збігається з розподілом для системи з абсолютним пріоритетом.

Розрахуємо розподіл ймовірностей станів для шару Γ_I (рис.3.11).

[illegible]

Звідси

$$P'_{1i_2} = P_{0i_2} \frac{\lambda_1}{\lambda_2 + \mu_2}.$$

Далі

$$\begin{aligned}\lambda_1 P_{11'} - P'_{21'} (\lambda_2 + \mu_2) &= 0, \\ \lambda_1 P_{12'} - P'_{22'} (\lambda_2 + \mu_2) &= 0, \\ &\dots\dots\dots\\ \lambda_1 P_{1n_2} - P'_{2n_2'} (\lambda_2 + \mu_2) &= 0.\end{aligned}$$

Звідси

$$P'_{2i_2'} = P'_{1i_2'} \frac{\lambda_1}{\lambda_2 + \mu_2} = P_{0i_2} \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^2.$$

$$\begin{aligned} \lambda_1 P_{n_1-1,1'} - P'_{n_2 1'} (\lambda_2 + \mu_2) &= 0, \\ \lambda_1 P_{n_1-1,2'} - P'_{n_2 2'} (\lambda_2 + \mu_2) &= 0, \\ &\dots\dots\dots \\ \lambda_1 P_{n_1-1,n_2} - P'_{n_1 n_2} (\lambda_2 + \mu_2) &= 0. \end{aligned}$$

Звідси

$$P'_{n_1 i_2'} = P'_{n_{1-1} i_2'} \frac{\lambda_1}{\lambda_2 + \mu_2} = P_{0 i_2} \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{n_2}.$$

Таким чином, умовні ймовірності станів для шару Γ_l визначаються співвідношенням

$$P_{i_1 i_2} = P_{0 i_2} \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{i_1}, \quad i_1 = 1, 2, \dots, n_1; \quad i_2 = 1, 2, \dots, n_2. \quad (3.49)$$

Тепер використовуємо співвідношення (3.18) для розрахунку інтенсивностей групових переходів. Оскільки

$$P'_{0i} = \frac{\frac{\rho_2^i (1 - \rho_1^{n_1+1})^{i-1}}{(1 - \rho_1)^{i-1}}}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^l}, \quad i = 1, 2, \dots, n_2,$$

тоді

$$P'_{i_1 i_2} = \frac{\left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{i_1} \frac{\rho_2^{i_2} (1 - \rho_1^{n_1+1})^{i_2-1}}{(1 - \rho_1)^{i_2-1}}}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{(1 - \rho_1)} \right]^l}.$$

З іншого боку

$$\begin{aligned} \lambda_{01} &= \lambda_1 \sum_{i=1}^{n_2} P_{0i}' = \lambda_1 \frac{\sum_{i=1}^{n_2} \frac{\rho_2^i (1 - \rho_1^{n_1+1})^{i-1}}{(1 - \rho_1)^{i-1}}}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{(1 - \rho_1)} \right]^l} = \lambda_1 \rho_2 \frac{\sum_{l=0}^{n_2-1} \left[\frac{\rho_2^i (1 - \rho_1^{n_1+1})^{i-1}}{1 - \rho_1} \right]^l}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^l} = \\ &= \lambda_1 \rho_2 \left[1 - \frac{\left(\frac{\rho_2^i (1 - \rho_1^{n_1+1})^{i-1}}{1 - \rho_1} \right)^{n_2}}{\sum_{l=0}^{n_2} \left[\frac{\rho_2 (1 - \rho_1^{n_1+1})}{1 - \rho_1} \right]^l} \right]. \end{aligned} \tag{3.50}$$

Далі

$$\lambda_{10} = \mu_2 \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} P_{i_1 i_2} = \mu_2 \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} P_{0 i_2} \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{i_2} = \mu_2 \sum_{i_1=1}^{n_1} \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{i_1} \sum_{i_2=1}^{n_2} P_{0 i_2}.$$

Оскільки

$$\sum_{i_2=0}^{n_2} P_{0 i_2} = \rho_2 \left[1 - \frac{\left(\frac{\rho_2 (1 - \rho_1^{n+1})}{1 - \rho_1} \right)^{n_2}}{\sum_{l=0}^{n_2} \left(\frac{\rho_2 (1 - \rho_1^{n+1})}{1 - \rho_1} \right)^l} \right],$$

тоді

$$\lambda_{10} = \mu_2 \rho_2 \left[1 - \frac{\left(\frac{\rho_2 (1 - \rho_1^{n+1})}{1 - \rho_1} \right)^{n_2}}{\sum_{l=0}^{n_2} \left(\frac{\rho_2 (1 - \rho_1^{n+1})}{1 - \rho_1} \right)^l} \right] \sum_{i_1=1}^{n_1} \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^i =$$

, (3.51)

$$= \lambda_2 B \frac{\frac{\lambda_1}{\lambda_2 + \mu_2} - \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{n_2}}{1 - \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)} = \lambda_2 B \frac{\lambda_1 - \left(\frac{\lambda_1}{\lambda_2 + \mu_2} \right)^{n_2} (\lambda_2 + \mu_2)}{\lambda_2 + \mu_2 - \lambda_1}$$

де

$$B = 1 - \frac{\left(\frac{\rho_2 (1 - \rho_1^{n+1})}{1 - \rho_1} \right)^{n_2}}{\sum_{i=0}^{n_2} \left(\frac{\rho_2 (1 - \rho_1^{n+1})}{1 - \rho_1} \right)^i}.$$

Тепер, об'єднуючи (3.48), (3.50), (3.51), (3.18), можливо записати остаточні вирази для ймовірностей станів системи з відносним пріоритетом.

Оцінимо ефективність обслуговування в системі з пріоритетом одних запитів перед іншими. Нехай в одноканальну систему надходить суперпозиція двох потоків. Проаналізуємо вплив пріоритету на середню довжину черги для кожного з потоків в системі з абсолютним пріоритетом одних запитів перед іншими.

Введемо

$$\eta_0 = \frac{\bar{n}_{20}}{n_{10}},$$

\bar{n}_{10} — середня довжина черги для першого потоку в системі без пріоритету,

\bar{n}_{20} — середня довжина черги для другого потоку в системі без пріоритету,

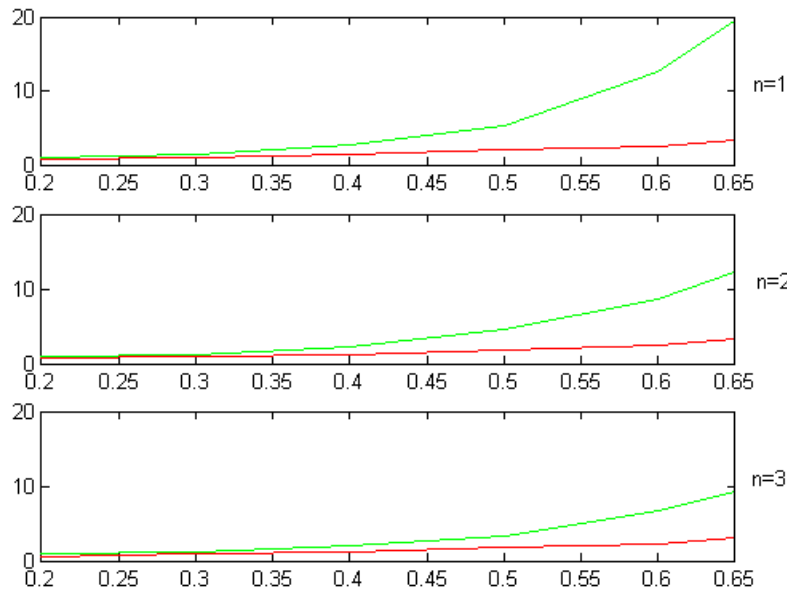
$$\eta_0 = \frac{\bar{n}_{21}}{n_{11}},$$

\bar{n}_{11} — середня довжина черги для першого потоку в системі з абсолютним пріоритетом,

\bar{n}_{21} — середня довжина черги для другого потоку в системі з абсолютним пріоритетом.

$$\alpha_1 = \frac{\lambda_1}{\mu_1}; \alpha_2 = \frac{\lambda_2}{\mu_2}; n = 1$$

Зафіксуємо α_2 , наприклад $\alpha_2 = 0,3n$, і проаналізуємо залежність η_0 і η_1 від α_1 . Відповідні графіки наведені на рис. 3.12.



$P_{відм_2}$ — — відношення середніх довжин черг в системі без пріоритетів,
 $P_{відм_2}^0$ — — відношення середніх довжин черг в системі з абсолютним пріоритетом першого потоку.

Рис. 3.12. Залежність довжини черги для одного потоку від інтенсивності іншого.

Аналіз графіків для $n = 1$ дозволяє зробити наступні висновки. В системі без пріоритету середня довжина черги для кожного з потоків при заданому числі каналів та інтенсивності обслуговування залежить від інтенсивності цього потоку і не залежить від інтенсивності іншого потоку. В системах з абсолютним пріоритетом справа йде не так. Середня довжина черги для потоку, що має абсолютний пріоритет, так само, як і раніше, залежить тільки від його інтенсивності. Навпаки, для другого потоку, який не володіє пріоритетом, запити якого обслуговуються тільки при відсутності запитів першого потоку, середня довжина черги визначається не тільки його інтенсивністю, але й інтенсивністю пріоритетного потоку. При цьому середня довжина черги для запитів цього потоку тим більше, чим вище інтенсивність пріоритетного потоку.

Зауважимо, що відповідна залежність є нелінійною, оскільки приводить до показового зростання довжини черги для запитів непріоритетного потоку при збільшенні інтенсивності пріоритетного. У багатьох практичних ситуаціях доцільно послабити істотну різницю, яка при цьому виникає, в ефективності обслуговування різних за пріоритетом потоків. Природна можливість поліпшення ситуації, що виникає у випадках, коли довжина черги

непріоритетних запитів наближається до гранично допустимої, полягає в знятті пріоритету на час, достатній для відновлення відносної рівноваги в довжинах черг. Інший шлях вирішення цього завдання при збереженні пріоритетної дискримінації, пов'язаний зі збільшенням числа каналів. Результати відповідних прорахунків для випадків, коли $n=2$ і $n=3$ приведені також на рис. 3.12. З наведених на цьому рисунку графіків видно, що при збільшенні кількості каналів (але тому самому їх завантаженні) перевага, створювана пріоритетом, знижується, і різниця в середніх довжинах черг стає менш суттєвою.

3.6. Наближена технологія оцінки ефективності багатоканальної системи з відмовами і неоднорідним вхідним потоком з пріоритетами

Нехай на вхід n -канальної СМО з відмовами надходить суперпозиція l потоків, що мають різні пріоритети. Введемо нумерацію потоків таким чином, щоб найбільш пріоритетний мав номер 1, наступний за пріоритетом – номер 2 і, нарешті, найменш пріоритетний – номер l . Нехай, далі, всі вхідні потоки є пуассонівськими, тривалості обслуговування запитів всіх типів розподілені експоненційно, причому інтенсивності потоку запитів k -го потоку дорівнює λ_k , а інтенсивності їх обслуговування μ_k , $k=1,2,\dots,l$. Зрозуміло, що в такій системі запити k -го потоку обслуговуються тільки в тому випадку, коли відсутні запити більш пріоритетних потоків з номерами $1,2,\dots,k-1$. Безпосередній аналіз ефективності такої системи, схожий до того, який був проведений для найпростішого окремого випадку суперпозиції двох потоків, у випадках, коли число каналів завелике, теоретично може бути виконано, але стає надзвичайно трудомістким завданням. Разом з тим, наближені оцінки ефективності можуть бути отримані з використанням наступних елементарних міркувань.

Запити найбільш пріоритетного першого типу обслуговуються в цій системі так, як ніби запитів інших типів просто не існує. Тому показники ефективності обслуговування для цих запитів визначаються формулами Ерланга, тобто ймовірність зайнятості дорівнює s_l каналів, ймовірність відмови запиту першого типу і середнє число зайнятих каналів, визначаються наступними співвідношеннями [208]

$$P_{s_1} = \frac{\frac{\left(\frac{\lambda_1}{\mu_1}\right)^{s_1}}{s_1!}}{\sum_{q=0}^n \frac{\left(\frac{\lambda_1}{\mu_1}\right)^q}{q!}}, \quad s_1 = 0, 1, 2, \dots, n. \quad (3.52)$$

$$P_{\text{відм}_1} = \frac{\frac{\left(\frac{\lambda_1}{\mu_1}\right)^n}{n!}}{\sum_{q=0}^n \frac{\left(\frac{\lambda_1}{\mu_1}\right)^q}{q!}}, \quad (3.53)$$

$$\bar{S}_1 = \frac{\lambda_1}{\mu_1} (1 - P_{\text{відм}_1}). \quad (3.54)$$

Тоді для обслуговування другого і наступних за пріоритетністю потоків в середньому вільні $n - \bar{S}_1$ каналів і, оскільки запити другого потоку володіють абсолютним пріоритетом перед запитами всіх інших потоків, то ефективність їх обслуговування також описується формулами Ерланга. При цьому ймовірність зайнятості обслуговуванням запитів другого типу дорівнює s_2 каналів, ймовірність відмови запиту цього типу і середнє число каналів, зайнятих обслуговуванням цих запитів, відповідно дорівнюють

$$P_{s_2} = \frac{\frac{\left(\frac{\lambda_2}{\mu_2}\right)^{s_2}}{s_2!}}{\sum_{q=0}^{n-\bar{S}_1} \frac{\left(\frac{\lambda_2}{\mu_2}\right)^q}{q!}}, \quad s_2 = 0, 1, 2, \dots, n - \bar{S}_1. \quad (3.55)$$

$$P_{\text{відм}_2} = \frac{\frac{\left(\frac{\lambda_2}{\mu_2}\right)^{n-\bar{S}_1}}{(n-\bar{S}_1)!}}{\sum_{q=0}^{n-\bar{S}_1} \frac{\left(\frac{\lambda_2}{\mu_2}\right)^{q_1}}{q!}}. \quad (3.56)$$

$$\bar{S}_2 = \frac{\lambda_2}{\mu_2} (1 - P_{\text{відм}_2}) = \frac{\lambda_2}{\mu_2} \left(1 - \frac{\frac{\left(\frac{\lambda_2}{\mu_2}\right)^{n-\bar{S}_1}}{(n-\bar{S}_1)!}}{\sum_{q=0}^{n-\bar{S}_1} \frac{\left(\frac{\lambda_2}{\mu_2}\right)^{q_1}}{q!}} \right). \quad (3.57)$$

Зрозуміло, що в цих співвідношеннях обчислене значення $n - \bar{S}_1$ середнього числа каналів, вільних для обслуговування другого і наступних за пріоритетом потоків, необхідно округляти до найближчого цілого числа, що, природно, вносить похибку в результати розрахунків.

Продовжуючи аналогічно, для запитів довільного k -го типу запишемо формули, що оцінюють ефективність обслуговування запитів цього типу

$$P_{s_k} = \frac{\frac{\left(\frac{\lambda_k}{\mu_k}\right)^{s_k}}{s_k!}}{\sum_{q=0}^{n-\sum_{r=1}^{k-1} \bar{S}_r} \frac{\left(\frac{\lambda_k}{\mu_k}\right)^q}{q!}}, \quad s_k = 0, 1, 2, \dots, n - \sum_{r=1}^{k-1} \bar{S}_r, \quad (3.58)$$

$$P_{\text{від}M_k} = \frac{\frac{\left(\frac{\lambda_k}{\mu_k}\right)^{n - \sum_{r=1}^{k-1} \bar{S}_r}}{\left(n - \sum_{r=1}^{k-1} \bar{S}_r\right)!}}{\sum_{q=0}^{n - \sum_{r=1}^{k-1} \bar{S}_r} \frac{\left(\frac{\lambda_k}{\mu_k}\right)^q}{q!}}, \quad (3.59)$$

$$\bar{S}_k = \frac{\lambda_k}{\mu_k} (1 - P_{\text{від}M_k}) = \frac{\lambda_k}{\mu_k} \left(1 - \frac{\frac{\left(\frac{\lambda_k}{\mu_k}\right)^{n - \sum_{r=1}^{k-1} \bar{S}_r}}{\left(n - \sum_{r=1}^{k-1} \bar{S}_r\right)!}}{\sum_{q=0}^{n - \sum_{r=1}^{k-1} \bar{S}_r} \frac{\left(\frac{\lambda_k}{\mu_k}\right)^q}{q!}} \right). \quad (3.60)$$

Оцінимо похибку, відповідну отриманим співвідношенням. З цією метою порівняймо ймовірність відмови, яка визначається ними, з точним значенням, отриманим в результаті безпосереднього аналізу. Порівняння проведемо для випадку, коли на вхід надходить суперпозиція двох потоків, перший з яких має абсолютний пріоритет перед другим.

Отримаємо точні співвідношення для розподілу ймовірностей станів такої системи. Граф станів і переходів системи має вигляд, представлений на рис. 3.13.

Для аналізу цієї системи знову використовуємо технологію фазового укрупнення станів. Введемо сукупність групових станів наступним чином

$$\begin{aligned} E_{\Gamma_0} &= \{E_{00}, E_{01}, \dots, E_{0n}\}, \\ E_{\Gamma_1} &= \{E_{10}, E_{11}, \dots, E_{1,n-1}\}, \\ &\dots\dots\dots \\ E_{\Gamma_{i_1}} &= \{E_{i_10}, E_{i_11}, \dots, E_{i_1, n-i_1}\}, \\ E_{\Gamma_{n-1}} &= \{E_{n-1,0}, E_{n-1,1}\}, \\ E_{\Gamma_n} &= \{E_{n,0}\}. \end{aligned}$$

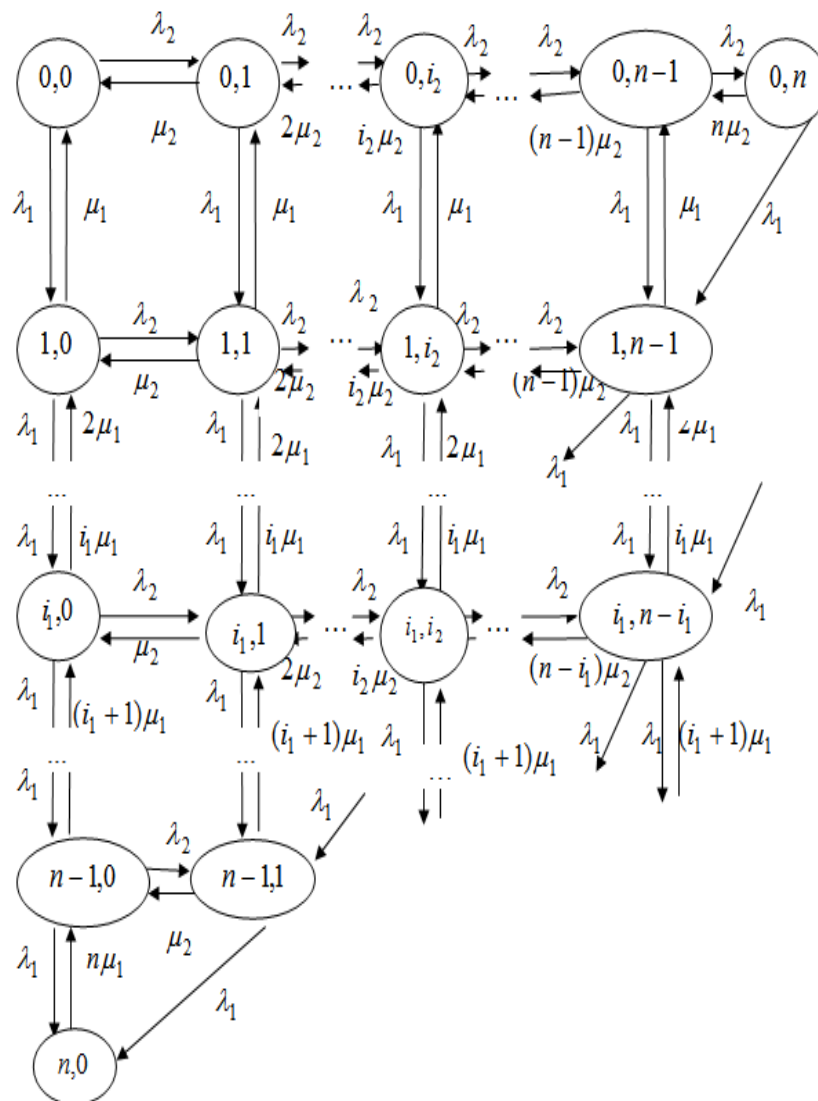


Рис. 3.13. Граф станів і переходів в n -канальній системі з відмовами, на вході якої є два потоки, перший потік має абсолютний пріоритет.

Граф станів і переходів укрупненої системи наведено на рис.3.14.

Розрахуємо умовні розподілу ймовірностей станів для кожного з шарів за умови перебування в цьому шарі.

З рисунка 3.13 зрозуміло, що шукані умовні ймовірності станів для шарів $E_{\Gamma_0}, E_{\Gamma_1}, \dots, E_{\Gamma_{n-1}}$ визначаються ерлангівськими розподілами для систем з відмовами і числом каналів, що дорівнюють відповідно $n, n-1, \dots, 2$.

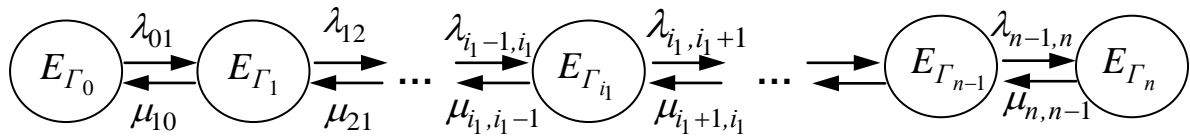


Рис. 3.14. Граф станів і переходів в укрупненої системі.

При цьому маємо

$$P_{0i_2} = \frac{\frac{\alpha_2^{i_2}}{i_2!}}{\sum_{l=0}^n \frac{\alpha_2^l}{l!}}, \quad i_2 = 0, 1, 2, \dots, n,$$

$$P_{1i_2} = \frac{\frac{\alpha_2^{i_2}}{i_2!}}{\sum_{l=0}^{n-1} \frac{\alpha_2^l}{l!}}, \quad i_2 = 0, 1, 2, \dots, n-1,$$

.....

$$P_{i_1 i_2} = \frac{\frac{\alpha_2^{i_2}}{i_2!}}{\sum_{l=0}^{n-i_1} \frac{\alpha_2^l}{l!}}, \quad i_2 = 0, 1, 2, \dots, n-i_1,$$

.....

$$P_{n-1, i_2} = \frac{\frac{\alpha_2^{i_2}}{i_2!}}{\sum_{l=0}^1 \frac{\alpha_2^l}{l!}}, \quad i_2 = 0, 1.$$

Тепер легко знайти інтенсивності переходів між шарами. Оскільки умовні розподіли ймовірностей станів в кожному з шарів задовольняють умові нормування, то

$$\lambda_{01} = \lambda_1 \sum_{i_2=0}^n P'_{0i_2} = \lambda_1,$$

$$\lambda_{01} = \lambda_1 \sum_{i_2=0}^n P'_{0i_2} = \lambda_1,$$

$$\lambda_{12} = \lambda_1 \sum_{i_2=0}^{n-1} P'_{1i_2} = \lambda_1,$$

.....

$$\lambda_{i,i_1+1} = \lambda_1 \sum_{i_2=0}^{n-i_1} P'_{ii_2} = \lambda_1,$$

.....

$$\lambda_{n-1,n} = \lambda_1 \sum_{i_2=0}^1 P'_{n-1,i_2} = \lambda_1,$$

$$\mu_{10} = \mu_1 \sum_{i_2=1}^{n-1} P'_{2i_2} = \mu_1,$$

$$\mu_{21} = 2\mu_1 \sum_{i_2=1}^{n-2} P'_{2i_2} = 2\mu_1,$$

.....

$$\mu_{i_1,i_1-1} = i_1 \mu_1 \sum_{i_2=1}^{n-i_1} P'_{ii_2} = i_1 \mu_1,$$

.....

$$\mu_{n,n-1} = n\mu_1.$$

Звідси зрозуміло, що система групових станів функціонує $E_{\Gamma_0}, E_{\Gamma_1}, \dots, E_{\Gamma_n}$ так само як n –канальна система з відмовами з вхідним потоком інтенсивністю λ_1 , запити якого обслуговуються з інтенсивністю μ_1 . Тоді розподіл ймовірностей групових станів має вигляд:

$$P_{\Gamma_{i_1}} = \frac{\frac{\alpha_1^{i_1}}{i_1!}}{\sum_{l=0}^n \frac{\alpha_1^l}{l!}}, \quad i_1 = 0, 1, 2, \dots, n.$$

Тепер легко записати співвідношення для безумовних ймовірностей станів системи:

$$P_{i_1 i_2} = P'_{i_1 i_2} \cdot P_{\Gamma_{i_1}} = \frac{\alpha_2^{n-i_1}}{\sum_{l=0}^{n-i_1} l!} \cdot \frac{\alpha_1^n}{\sum_{l=0}^n l!}, \quad i_1 = 0, 1, 2, \dots, n; \quad i_2 = 0, 1, 2, \dots, n; \quad i_1 + i_2 \leq n.$$

При оцінці ефективності такої системи цікавість представляють ймовірності відмов для запитів першого і другого потоків. Ясно, що відмова пріоритетного запиту з першого потоку відбудеться тільки в тому випадку, якщо всі n каналів системи зайняті обслуговуванням запитів тільки цього потоку.

Відповідний стан системи є $E_{n,0}$, а його ймовірність дорівнює

$$P_{\text{відм}_2} = P_{n,0} = \frac{\alpha_1^n}{\sum_{l=0}^n \frac{n!}{l!}}.$$

В результаті, запити пріоритетного потоку обслуговуються таким чином, ніби другого потоку взагалі не існувало.

З іншого боку відмова запиту з другого потоку настає у всіх випадках, коли всі n каналів системи зайняті обслуговуванням будь-яких запитів. Таким чином, відмовними для другого потоку будуть стани $E_{0,n}, E_{1,n-1}, \dots, E_{i_1, n-i_1}, \dots, E_{n-1,1}, E_{n,0}$. Ймовірність того, що система виявиться в одному з цих станів, дорівнює

$$P_{\text{відм}_2}^0 = \sum_{l=0}^n P_{i_1, n-i_1} = \sum_{i_1=0}^n \left(\frac{\frac{\alpha_1^{i_1}}{i_1!} \cdot \frac{\alpha_2^{n-i_1}}{(n-i_1)!}}{\frac{\sum_{l=0}^n \alpha_1^l}{l!} \cdot \frac{\sum_{l=0}^n \alpha_2^l}{l!}} \right) = \frac{1}{\sum_{l=0}^n \frac{\alpha_1^l}{l!}} \sum_{i_1=0}^n \frac{\alpha_1^{i_1} \cdot \alpha_2^{n-i_1}}{i_1! \cdot (n-i_1)!}. \quad (3.61)$$

Розрахуємо тепер похибку, що дається співвідношенням (3.56). Запишемо це співвідношення детально

$$P_{відм_2} = \frac{\left(\frac{\lambda_2}{\mu_2} \right) \left[n - \frac{\lambda_1}{\mu_1} \left(1 - \frac{\left(\frac{\lambda_1}{\mu_1} \right)^n}{\sum_{q=0}^n \frac{\left(\frac{\lambda_1}{\mu_1} \right)^q}{q!}} \right) \right]}{\left[n - \frac{\lambda_1}{\mu_1} \left(1 - \frac{\left(\frac{\lambda_1}{\mu_1} \right)^n}{\sum_{q=0}^n \frac{\left(\frac{\lambda_1}{\mu_1} \right)^q}{q!}} \right) \right]!} \cdot \frac{\sum_{q=0}^n \frac{\left(\frac{\lambda_2}{\mu_2} \right)^q}{q!}}{\left(\frac{\lambda_2}{\mu_2} \right)^q} . \quad (3.62)$$

Тут $[]$ – символ округлення до найближчого цілого.

Нехай наведені щільності для першого і другого потоків рівні поміж собою, тобто $\frac{\lambda_1}{\mu_1} = \frac{\lambda_2}{\mu_2} = \alpha$. Тоді співвідношення (3.62) набуває дещо простіший вигляд:

$$P_{widm_2} = \frac{\alpha^{-\left[n-\alpha \left(1-\frac{\frac{\alpha^n}{n!}}{\sum_{q=0}^n \frac{\alpha^q}{q!}} \right) \right]}}{\left[n-\alpha \left(1-\frac{\frac{\alpha^n}{n!}}{\sum_{q=0}^n \frac{\alpha^q}{q!}} \right) \right]! \sum_{q=0}^n \left[n-\alpha \left(1-\frac{\frac{\alpha^n}{n!}}{\sum_{q=0}^n \frac{\alpha^q}{q!}} \right) \right] \frac{\alpha^q}{q!}} . \quad (3.63)$$

Спроститься й співвідношення (3.61):

$$P_{widm_2}^0 = \frac{1}{\sum_{l=0}^n \frac{\alpha^l}{l!}} \sum_{i_1=0}^n \frac{\frac{\alpha^n}{i_1! (n-i_1)!}}{\sum_{l=0}^{n-i_1} \frac{\alpha^l}{l!}} \quad (3.64)$$

Побудуємо тепер систему графіків, що відображають залежності ймовірності P_{widm_2} від величини наведеної щільності потоків запитів, які розраховуються за формулами (3.63) і (3.64) для різного числа каналів.

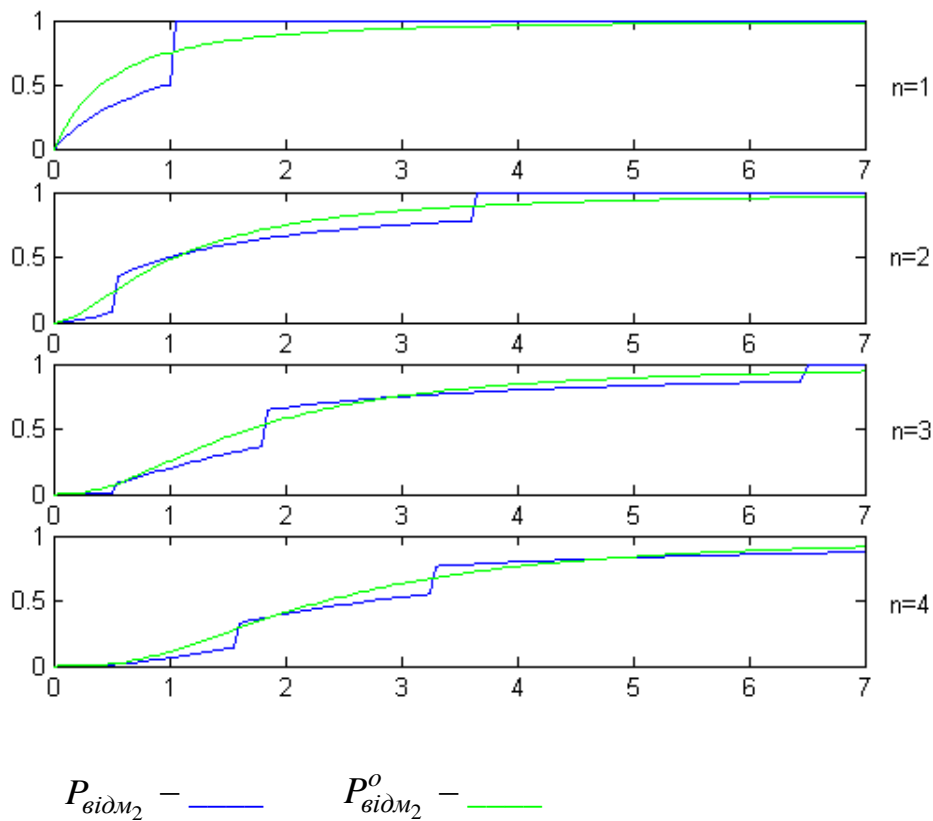


Рис. 3.15. Залежності ймовірностей P_{vidM_2} , $P_{vidM_2}^0$ від щільності потоків запитів, які розраховано для різного числа каналів.

Проведемо аналіз похибки співвідношення (3.64), використовуючи графіки, наведені на рис. 3.15 і таблицях 3.1 і 3.2.

Проведених розрахунків досить, щоб встановити загальну тенденцію: похибка співвідношення (3.63) невелика (9%) і спадає зі зростанням α і n .

Зауважимо, що і цю похибку, основна частка якої пов'язана з операцією округлення, можна знизити, якщо при обчисленні факторіала в (3.56) — (3.60) використовувати гамма-функцію

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt .$$

Таблиця 3.1 – Залежність ймовірності $P_{відм_2}$ від щільності потоків запитів, розрахованих для різного числа каналів

n	α					
	1	2	3	4	5	6
6	0,0030675	0,095238	0,34615	0,61538	0,67568	0,72
7	0,00051099	0,036697	0,20611	0,4507	0,52966	0,72
8	7,2993e—005	0,012085	0,11005	0,31068	0,52966	0,59016
9	9,124e—006	0,0034409	0,052157	0,19907	0,39834	0,59016
10	1,0138e—006	0,00085948	0,021864	0,11716	0,28487	0,46957
11	1,0138e—007	0,00019096	0,0081324	0,062749	0,19185	0,3604
12	9,2162e—009	3,819e—005	0,0027035	0,03042	0,12052	0,26492

Таблиця 3.2 – Залежність ймовірності $P_{відм_2}^0$ від щільності потоків запитів, які розраховано для різного числа каналів

n	α					
	1	2	3	4	5	6
6	0,013089	0,14819	0,37015	0,56195	0,6951	0,78229
7	0,003622	0,075825	0,25368	0,44972	0,60577	0,71558
8	0,00088751	0,035102	0,16115	0,34116	0,50931	0,63925
9	0,00019478	0,01479	0,094973	0,24464	0,41135	0,55544
10	3,8656e—005	0,0057036	0,052032	0,16567	0,31824	0,46779
11	6,9951e—006	0,0020231	0,026557	0,10595	0,23545	0,38087
12	1,1625e—006	0,00066316	0,012655	0,064015	0,16645	0,29928

При цьому, зокрема, співвідношення (3.63) прийме вигляд

$$P_{\text{від}M_2} = \frac{\alpha \left[n - \alpha \left(1 - \frac{\frac{\alpha^n}{n!}}{\sum_{q=0}^n \frac{\alpha^q}{q!}} \right) \right]}{\Gamma \left[n - \alpha \left(1 - \frac{\frac{\alpha^n}{n!}}{\sum_{q=0}^n \frac{\alpha^q}{q!}} \right) \right]}$$

Відповідні графіки показані на рис. 3.16.

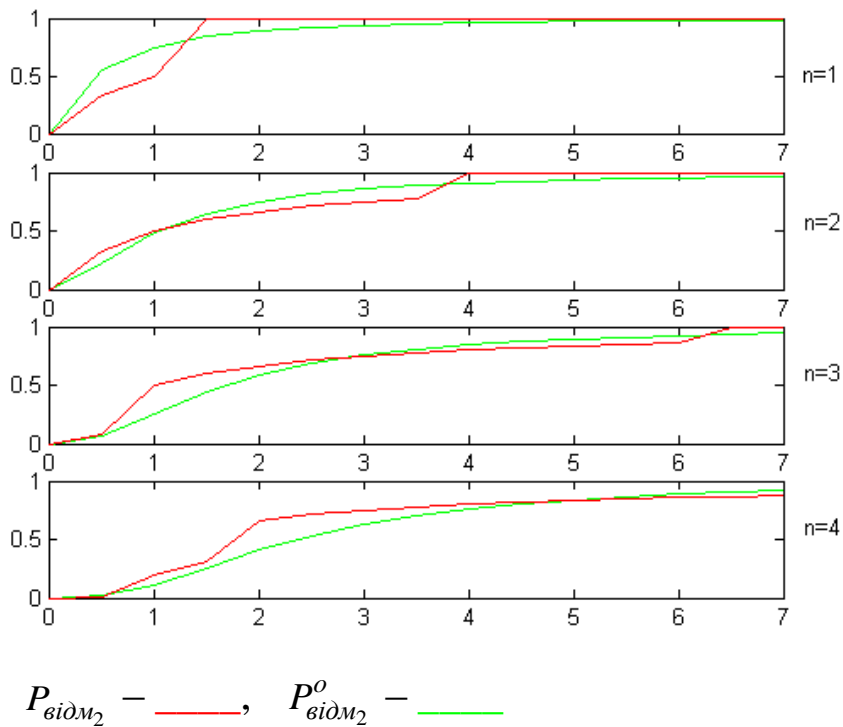


Рис. 3.16. Залежності ймовірностей $P_{\text{від}M_2}$, $P_{\text{від}M_2}^0$ від щільності потоків запитів, розрахованих для різного числа каналів.

Результат показує, що в системі з двома вхідними потоками, точність оцінки ефективності обслуговування менш пріоритетних запитів з урахуванням

середнього числа каналів, зайнятих обслуговуванням потоку більш пріоритетних запитів, є прийнятною. Це ж міркування може бути використано при оцінці точності розрахунку ефективності обслуговування запитів в системі з трьома потоками з урахуванням середнього числа каналів, зайнятих обслуговуванням більш пріоритетних каналів.

Таким чином, отримана технологія наближеної оцінки ефективності функціонування n -канальної СМО, на вхід якої надходить суперпозиція потоків з різним абсолютним пріоритетом. Простота цієї технології, а також прийнятна точність забезпечують можливість її практичного використання.

Висновки.

Система управління інформаційною технологією являє собою розподілене середовище, яке забезпечує інтеграцію всіх рівнів інформаційної системи в єдину систему, забезпечуючи управління інформаційною технологією будь-якої складності, що включає в себе безліч операційних систем, додатків, розподілені мережі і мережний сервіс, дозволяє швидко адаптувати інформаційну систему до поточних потреб завдань, що вирішуються при реалізації бізнес-процесів.

Відомі моделі функціонування інфокомунікаційних мереж використовують марківські моделі СМО. Отримані тут результати великі і різноманітні. Розглянуті та вирішені повністю завдання аналізу одно- і багатоканальних систем, на вхід яких надходить однорідний потік запитів або суперпозиція потоків без пріоритетів та з пріоритетами. При цьому отримані закони розподілу ймовірностей станів системи.

В розділі запропоновано прості практичні методики аналізу багатовхідних і багатоканальних інфокомунікаційних мереж. Методики реалізують технології декомпозиції складної задачі до сукупності більш простих підзавдань. Описані методики можуть бути використані для аналізу будь-яких марківських систем високої розмірності. Подальші дослідження можуть бути спрямовані на побудову ієрархічних процедур фазового укрупнення, необхідність яких виникне при дослідженні систем надвисокої розмірності.

Контрольні питання.

1. Побудуйте схему функціонування вузла мережі із неоднорідним потоком пакетів з точки зору теорії масового обслуговування.

2. Побудуйте граф станів та переходів одноканальної системи обслуговування пакетів для ситуації, коли на вхід вузла надається суперпозиція потоків пакетів, яка складається із двох потоків без пріоритетів.

3. Побудуйте граф станів та переходів одноканальної системи обслуговування пакетів для ситуації, коли на вхід вузла надається три потоки пакетів без пріоритетів.

4. Побудуйте граф станів та переходів двоканальної системи обслуговування пакетів для ситуації, коли на вхід вузла надається два потоки пакетів без пріоритетів.

5. Дайте перелік видів пріоритетів пакетів. Дайте пояснення кожному виду.

6. Побудуйте граф станів та переходів одноканальної системи обслуговування пакетів для ситуації, коли на вхід вузла надається два потоки пакетів, причому перший потік має абсолютний пріоритет перед другим.

7. Побудуйте граф станів та переходів одноканальної системи обслуговування пакетів для ситуації, коли на вхід вузла надається два потоки пакетів, причому перший потік має відносний пріоритет перед другим.

8. Побудуйте граф станів та переходів одноканальної системи обслуговування пакетів для ситуації із груповим вхідним потоком.

9. Побудуйте граф станів та переходів системи із повторною передачею загублених пакетів.

10. Обґрунтуйте принцип фазового укрупнення станів марківської системи надвисокої розмірності.

ВИСНОВКИ

У посібнику показано тенденції розвитку теорії телетрафіку для систем, які функціонують в умовах реальних потоків вимог і процесу обслуговування, за рахунок розширення наукової бази моделей і методів аналізу та прогнозування трафіку, який представляє собою суперпозицію немарківських потоків, в багатоканальних вузлах телекомунікаційних мереж з метою підвищення загальносистемних показників якості обслуговування QoS (затримка пакета у вузлі, рівень втрат пакетів).

Розвиток теорії телетрафіка показано на комплексі розроблених і вдосконалених одно— та багатоканальних моделей нестационарного та неоднорідного трафіку, моделей багатоканальних вузлів мереж з пріоритетами потоків. Дано обґрунтування абсолютного та відносного пріоритетів.

У роботі розглянуто метод оцінки ефективності функціонування телекомунікаційної мережі з неординарним груповим вхідним потоком. Отримано співвідношення для розрахунків розподілу ймовірностей станів системи для довільного розподілу кількості пакетів в групі. При цьому показано, що для одноканальної системи передачі ймовірність обслуговування не залежить від характеру розподілу кількості пакетів в групі, а визначається тільки його середнім значенням.

У роботі запропоновано метод аналізу марківської системи високої розмірності з використанням технології фазового укрупнення станів і вдосконалений за рахунок більш обґрунтованого вибору початкового розподілу ймовірностей станів, а також використання більш ефективної процедури розрахунків проміжних розподілів ймовірностей. Показано, що метод фазового укрупнення станів марківських моделей вузлів мереж дозволяє істотно знизити розмірність завдання шляхом її декомпозиції.

Для марківського ланцюга надвисокої розмірності запропонована ієрархічна процедура фазового укрупнення станів, реалізована в ситуації, коли процес переходів в системі управляється марківським процесом.

Побудовано математичні моделі вузлів маршрутизації: для інтерфейсу вузла мережі з неоднорідним вхідним потоком пакетів без пріоритетів, інтерфейсу вузла мережі з неоднорідним вхідним потоком пакетів з абсолютним пріоритетом, багатоканального вузла маршрутизації з неоднорідним вхідним потоком пакетів з абсолютним пріоритетом, вузла мережі з неоднорідним вхідним потоком пакетів з відносним пріоритетом, вузла мережі з повторною передачею пакетів. Розроблена наближена технологія

оцінки якості обслуговування багатоканальних вузлів маршрутизації з відмовами і неоднорідним вхідним трафіком з пріоритетами.

Список використаних джерел

1. Таха, А. Х. Введение в исследование операций: пер. с англ. – 7-е издание/ Хемди А. Таха. –Москва: Издательский дом "Вильямс", 2005. – 912 с.
2. Мину, М. Математическое программирование / М. Мину – М.: Наука, 1990. – 488с.
3. Карманов, В.Г. Математическое программирование: учебное пособие / В.Г. Карманов – М.: Наука, 1980. – 256с.
4. Хедли, Дж. Нелинейное и динамическое программирование / Дж. Хедли – М.: Мир, 1967. – 561с.
5. Аттетков, А.В. Методы оптимизации: учеб. для вузов / Под ред. В.С. Зарубина, А. П. Крищенко. – 2-е изд., стереотип. – М.: МГТУ им. Н. Э. Баумана, 2003. – 440 с.
6. Логистика: Учеб. пособие / Под ред. Б.А. Аникина. – М.: ИНФРАМ–М, 1999. – 327 с. – ISBN 5-86225-958-9
7. Практикум по логистике: Учеб. пособие. — 2-е изд., перераб. и доп. / Под ред. Б.А.Аникина. — М.: ИНФРА–М, 2006.— 276 с.
8. Корпоративная логистика. 300 ответов на вопросы профессионалов / Под общ. и научн. редакцией проф. В.И. Сергеева. – М.: ИНФРА–М, 2005. - 976 с. – ISBN 5-16-001675-9
9. Мешкова, Л.Л. Логистика в сфере материальных услуг.: 2-е изд. / Л.Л. Мешкова, И.И. Белоус, Н.М. Фролов – Тамбов: Изд-во Тамб. гос. техн. ун-та, 2002. – 188с.
10. Основы логистики: Учеб. пособие/ Под ред. Л. Б. Миротина и В. И. Сергеева. – М.: ИНФРА–М, 2000. — 200 с. – ISBN 5-16-000003-8
11. Неруш, Ю. М. Логистика: учеб. — 4-е изд., перераб. и доп. /Ю.М. Неруш – М.: ТК Велби, Изд-во «Проспект», 2006. – 520 с.
12. Сергеев, В.И. Логистика: Информационные системы и технологии: учебно–практическое пособие. / В.И. Сергеев, М.Н. Григорьев, С.А. Уваров – М.: Издательство «Альфа–Пресс», 2008. – 608с.
13. Кобзарь, А.И. Прикладная математическая статистика / А.И. Кобзарь – М.: Физматлит, 2006. — 816 с.

Навчальне видання

ПУСТОВОЙТОВ Павло Євгенович

Математичні моделі мереж зв'язку

Навчальний посібник

з дисципліни

«Математичні моделі мереж зв'язку»

Для студентів спеціальності

172 Телекомунікації та радіотехніка

Відповідальний випусковий В.М. Безрук

Редактор

Комп'ютерна верстка

Українською мовою

В авторській редакції

План 2019 р. (друге півріччя), поз.

Підп. до друку __.__.2019. Формат 60x84 1/16

Умов. друк. Арк. 6,5

Зам. №

Облік. вид арк.7,7

Ціна договірна

Спосіб друку—ризотрафія

Тираж 100 прим.

ХНУРЕ, 61166, Харків, проспект Науки, 14

Віддруковано у редакційно—видавничому відділі ХНУРЕ

61166, Харків, проспект Науки, 14